

The trials of xenotransplantation

While some researchers report progress towards the goal of producing pig organs for human transplantation, others have revealed new causes for worry about the potential consequences.

The stark dilemma over the wisdom of proceeding with clinical trials of animal organ transplants in humans is crystallized in two papers that, in the interests of public debate, are being released this week by *Nature* in electronic form before their publication in the journal. One, by scientists at the Scottish company PPL Therapeutics, shows for the first time the successful cloning of pigs from adult somatic cells; a similar paper based on fetal cells is expected to be published in this week's issue of *Science*. The breakthrough opens the way to creating genetically modified animals with great precision. Knock out the α -1,3-galactosyl transferase gene, for example, and the pigs would lack the galactose sugar on their cell surfaces that is the principal cause of acute rejection of pig organs in humans.

But in another paper to appear in the same issue, Daniel Salomon from the Scripps Research Institute, and colleagues elsewhere, show that porcine endogenous retroviruses (PERV) can infect human cells in culture. Moreover, they go on to demonstrate that transplanting pig pancreatic islets into immunosuppressed mice leads to widespread infection with PERV.

Guidelines issued by the US Food and Drug Administration (FDA) in 1996 would have allowed weakly regulated xenotransplant trials. The agency has now toughened the guidelines considerably, pulling trials under federal (rather than local) control, for example, and placing a de facto ban on transplants from non-human primates (see *Nature* **405**, 606–607; 2000).

Contagious viruses are a major worry in xenotransplantation, as they carry the risk of creating man-made pandemics. But unlike HIV and other retroviruses, PERV seem unlikely to jump from

the recipient to others. And the work by Salomon *et al.* provides an animal model for basic research on viral transmission in xenotransplantation. This is likely to be followed by experiments with transplants into immunosuppressed monkeys.

Salomon points out that the techniques for studying PERV infection developed in the mouse model would also allow better monitoring for such infection in clinical trials. The FDA, the world's foremost regulatory health authority, is responsible for ensuring that trials are carried out safely. It is disconcerting that even this agency was, just four years ago, willing to endorse xenotransplantation trials of pig and other animal parts with little supervision. Pushing ahead with even limited trials could pave the way for widespread trials, perhaps in countries lacking authorities as competent as the FDA.

Xenotransplantation falls into the category of hazards where, although the risk is probably low — and the benefit to individuals undoubtedly substantial — the public consequences could be catastrophic. Indeed, Salomon's research also shows how little is known about PERV, while more dangerous viruses may also be present in pigs but remain undetected. The pathogenicity of animal viruses can also change unpredictably when they jump the species barrier. But there is virtually no research aimed at detecting and understanding such risks.

The FDA now requires trial sponsors to carry out a battery of tests, and all patients are monitored for infection. But more independent research on potential viral risks is needed. In risk assessment, absence of evidence is too often read as evidence of absence. What might we learn in four years that we do not know now? That is the case for a moratorium on clinical trials. ■

Seeing the wood for the trees

The spectacular US forest fires should be a stimulus to improving the management of the ecosystems in the American west.

This summer's news that 70 major fires are burning out of control in the western United States is a familiar story that gets increasingly worse with each passing year. Four million acres have been lost to fire so far this year. The construction of more suburban housing in the forest, prohibitions on logging that prevent the clearance of menacing brush, and the suppression for many decades of 'natural' forest burns, have each steadily raised the prospects for disaster in the region.

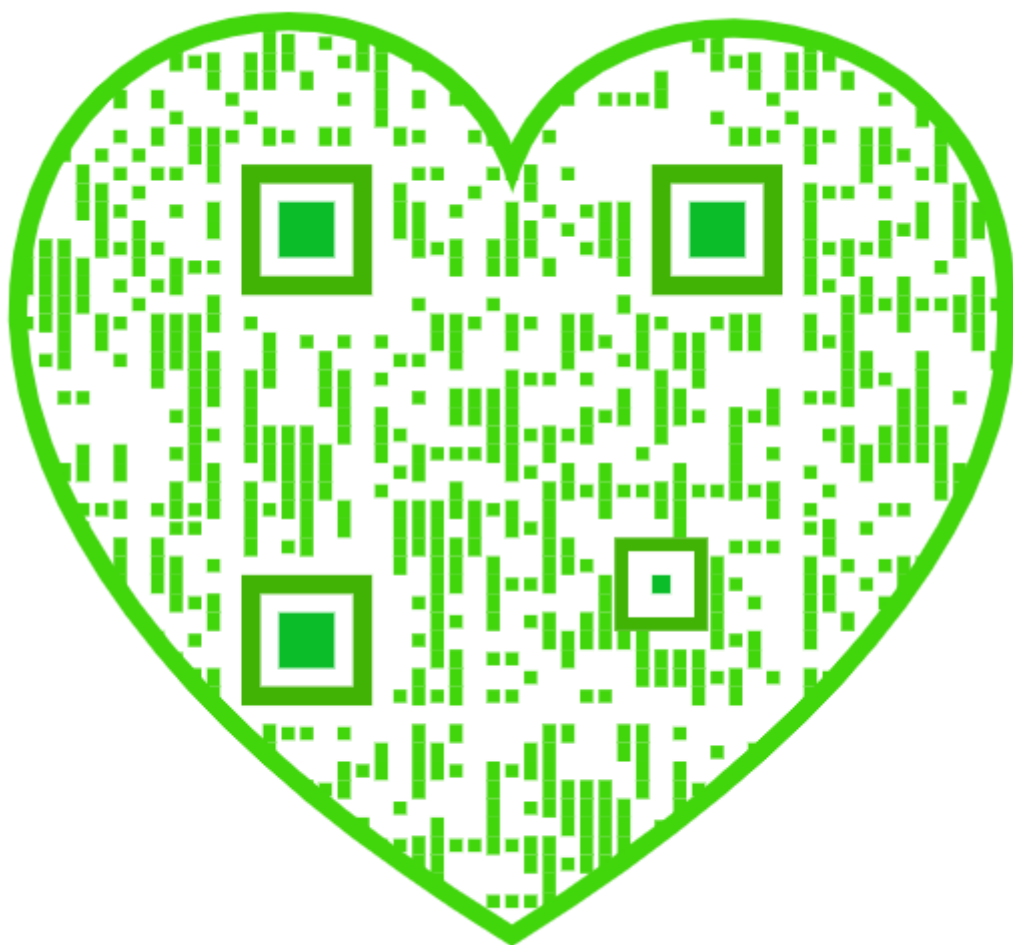
It would be wrong for ecologists or policymakers to pretend that additional research, improved forest management or even more thorough planning of land use can remove these prospects. The natural propensity of the region's forests to burn and the increasingly dispersed human population among these same forests are on a collision course, whatever steps are taken. But that doesn't mean that ecological science cannot assist policymakers in seeking to mitigate the problem. Steps could readily be taken to introduce sound ecological research and apply it to the situation, while dampening some of the passions that can obstruct sensible land-use policy.

The Forest Service, which is responsible for managing forest on federal land, is a fragmented and conservative organization, lacking strong scientific leadership and often beholden to local political

interests. It uses its money to pay its staff, not to conduct high-quality research into forest management. The Joint Fire Science Program, established by the Congress in 1998, is a welcome attempt to implement a more systematic approach to forestry research.

The region's frontier culture tends to value individual property rights more highly than the common interest, and instinctively rejects any form of government interference in planning, building or land use. Political leadership is therefore needed to resolve the dangerous and futile stand-off between the region's landowners and the federal government. But the determined efforts of Bruce Babbitt, the interior secretary in the Clinton administration, to address these issues have unfortunately been mistrusted by many in the western United States. His proposal for a National Biological Survey to map the nation's ecosystems, for example, was destroyed after a highly partisan battle.

This summer's fires must stimulate the federal government and state governments in the American west to break this impasse and address the issue of forest management directly. The problem is an opportunity for the new administration in Washington, of whichever stripe. It should appoint a bipartisan commission to design a new relationship between federal agencies such as the Forest Service, local government, land-use planning and ecosystem research. ■





Gene patents
German minister
brings rules into line
with rest of Europe
p664



Big ears
Astronomers plan
biggest-ever
radiotelescope
p665



Mars mission
NASA scales up
again with plans
for Mars rovers
p666



Animal welfare
US government
seeks revised
definition of distress
p668

German scientists left in the cold as Berlin rejects rescue plan

Alison Abbott and Ute Gitschel, Munich

At the time of German reunification in 1990, 1,500 hand-picked scientists from institutes run by the former east German Academy of Sciences were promised permanent jobs in universities. Initially, they were all supported by an interim programme, the Wissenschaftler Integrationsprogramm (WIP), and for a very small number this led to full-time employment.

But the majority have not been so lucky, and now 200 such scientists in Berlin face unemployment because the city is backing out of an agreement to co-finance a follow-up rescue programme. The other five *Länder* (states) continue to support the programme. And at a meeting last week, the so-called *WIPianers* agreed to lobby parliamentarians in Berlin to vote against the Berlin government's move to drop out of the programme.

They are threatening firmer action if their requests are rejected. "Some said that they'd be prepared to take harsher protest measures if no solution is found," says Gottfried Seifert, a retired mathematician from the Technical University in Berlin and spokesman for the Berlin *WIPianers*.

The *WIPianers* were selected by Germany's science council, the Wissenschaftsrat, as high-calibre scientists who could help build up the research capacities of east German universities. But in practice few were integrated into the universities, as the programme coincided with drastic cuts to their faculties (see *Nature* 378, 656–657; 1995).

Berlin, a powerful centre for science before reunification, was host to one-third of all *WIPianers*. So far only 16 have permanent faculty positions in Berlin universities.

When the WIP programme ended in 1996, the priority shifted from integrating the *WIPianers* to keeping them employed on short-term grants. A special programme was set up to achieve this, co-financed by the federal and *Länder* governments.

This programme finishes at the end of the year, and remaining *WIPianers* had welcomed a move by the federal research ministry to set up another joint programme, worth around DM150 million (US\$70 mil-



Sobotka: worried his team will be axed.

lion), to which they could apply.

Berlin, along with the five new *Länder*, signed up to the programme last year, committing itself to provide DM19 million in return for matching federal funds. But last month Berlin's science ministry said worsening financial circumstances meant that it could not provide matching funds.

Instead, the city has proposed that the money for co-financing any project under the programme should come from the universities, which are fully financed by the *Länder* government and where half of its 200 remaining *WIPianers* work. Non-university institutes, at which most of the remaining Berlin *WIPianers* work, would be legally barred from co-financing such projects as they are partly financed by the federal government.

But the universities, whose finances have

almost halved in the past ten years (see *Nature* 380, 278; 1996), say they cannot afford to do this. The Humboldt University, for example, is already selling some of its buildings to repay extended credit from the ministry. And Berlin's Technical University says that its 2001 budget is not even enough to pay the salaries of its existing staff.

The other five *Länder* are all prepared to co-finance the programme, both to help the *WIPianers* but also for more pragmatic reasons. In the state of Brandenburg, for example, 115 *WIPianers* are working in universities and research institutes on money from the special university programme.

Friedrich Buttler, state secretary for Brandenburg's research ministry, says the scientists bring in around the same amount of money in research grants as the programme costs to run. "They are running 88 different research projects, involving 230 scientists: this is a very good return on investment," he says. Buttler is enthusiastic about the successor programme.

The *WIPianers* point out that, in contrast,

Roslin backs off pig organ work

Declan Butler

The Roslin Institute in Edinburgh — where Dolly the sheep was cloned — has been working to create transgenic pigs as a source of organs for transplant to humans. But the future of such efforts appeared in doubt this week, following comments by researcher Ian Wilmut that Geron Bio-Med intended to cut back on the research.

Wilmut said that the California-based company — which was created by the purchase of the Roslin Institute's commercial arm Roslin Bio-Med by the US biotech company Geron (see *Nature* 399, 92; 1999) — was concerned about the risks that xenotransplantation might create viral diseases in humans.

But Grahame Bulfield, director of the Roslin, told *Nature* on Monday that Wilmut



Some fear that pig viruses could jump species.

was speaking in a personal capacity. In comments later confirmed by Geron Bio-Med he added that if the company were to abandon xenotransplantation work, it would be purely on the grounds of refocusing research priorities, and not viral risk.

Berlin stands to lose out on DM19 million in federal money if it refuses to co-finance the successor programme. "It is disgraceful," says Seifert, adding that during the three years of the special university programme, Berlin *WIPianers* won DM20 million in grant money.

The programme cost Berlin DM25 million, which was matched by federal funds. "Every mark paid by the city was converted to nearly three marks for research in Berlin," Seifert says.

But the average age of the remaining *WIPianers* is increasing and many fear that the excuse of impending retirement will be used to deal with the problem. Joachim Sobotka, a 62-year-old agricultural ecologist, says that if he is forced to retire early because his funds are discontinued, it will mean unemployment for the research team he has built up at Humboldt University, which has published 25 papers in the past few years.

Most Berlin *WIPianers* are unwilling to comment publicly on their plight, fearing that it could further jeopardize their employment prospects. But they plan to lobby Berlin's parliamentarians individually to ask them to vote against the science ministry's plans to withdraw from the new programme.

► <http://www.gew-berlin.de/wip/chronik1.htm>

German government takes a narrow view of gene patents

Quirin Schiermeier, Munich

Germany's research minister Edelgard Bulmahn promised last week that new legislation on patents on genetic material would specify that they should be given a narrow interpretation.

The pledge came at a meeting attended by Bulmahn and the minister for legal affairs, Herta Däubler-Gmelin, as well as academics, industrialists and patent experts. The meeting discussed changes to Germany's patent laws that are to bring them in line with the European directive on the legal protection of biotechnology.

The commitment to the new legislation — which should have been approved by the end of last month — was welcomed by representatives from the biotechnology industry. "Harmonized European patent rules will give us the legal security we need," says Thomas von Rüden, chief scientific officer of the Munich-based company MorphoSys.

But there was also agreement at the meeting that patents issued should be interpreted in a "narrow way". This reflects concerns expressed last month by German scientists that broad gene patents could hinder the



Bulmahn: wants laws friendly to research.

exploitation of newly discovered functions for DNA sequences (see *Nature* **406**, 111; 2000).

Bulmahn said that the legal comments that would accompany the directive should encourage a 'research-friendly' interpretation by national and European courts. Although not legally binding, these comments are aimed at guiding the interpretation of a law by the courts.

Some scientists oppose the idea that patents involving genetic information should get special treatment, believing that patents should be seen as incentives, rather than obstacles, to researchers.

But Detlev Ganten, scientific director of the Berlin-based Max Delbrück Center for Molecular Medicine, questions whether Bulmahn's concession goes far enough. "Gene patents should, on principle, be restricted to identified functions," he says.

But Detlev Ganten, scientific director of the Berlin-based Max Delbrück Center for Molecular Medicine, questions whether Bulmahn's concession goes far enough. "Gene patents should, on principle, be restricted to identified functions," he says.

Deal on reprints could mean royalties for scientists

Rex Dalton, San Diego

Many published scientists could be eligible for a share in a \$7.25 million settlement resulting from a federal lawsuit in Oakland, California. Approved by the US District Court late last month, the fund will compensate authors who have had reprints of published articles sold without their consent.

The class action lawsuit was launched two years ago when five authors sued UnCover, a document delivery firm based in Denver, Colorado, claiming that they had been cheated of royalty payments. Although the original plaintiffs were freelance writers and journalists, scientists are among the potential beneficiaries, as many of the articles distributed by UnCover were from scientific journals.

The compensation could be worth as much as \$30,000 per document delivery. But the actual amount will depend on the number of authors who successfully file claims, and may not be so high.

Any author who retained the rights to their work after initial publication is eligible to apply for compensation. Three-quarters of the settlement fund, less the estimated



\$3 million in legal costs for the plaintiffs' attorneys, has been set aside for such claims. The remaining quarter will go to authors who did not register their US copyright, but retained the right after first publication.

"Selling individual articles electronically without authors' permission has been an industry-wide practice," says John Shuff, one of the authors' attorneys. He expects the case to have a significant impact on the burgeoning world of Internet publishing.

In some cases, UnCover had made arrangements to secure rights from publishers to distribute documents. But it had not always reached agreement with the authors. As part of the settlement, UnCover — now owned by the British firm Ingenta — has agreed to seek the permission of authors and pay royalties in the future.

Robert Eisenbach, an attorney representing UnCover and Knight-Ridder, a previous owner of the company, said the settlement was reached to avoid future litigation costs.

Some observers believe that the agreement might hinder the distribution of scientific literature. Stevan Harnad, a professor of cognitive sciences at the University of Southampton in England, calls it "nothing but short-sighted nonsense". Others feel scientists are entitled to be paid for their work, just like any other author.

It is uncertain how many scientists are eligible for compensation because of the different arrangements between authors and publishers. Authors can check on the web whether UnCover distributed their articles. Claims must be filed by 27 October 2000.

► <http://www.uncoversettlement.com>



Sequencing solutions: but regulatory guidelines proposed so far have not been coordinated.

Japan seeks to unify ethics rules on genomics research

Robert Triendl, Tokyo

The Japanese government announced last week that it has created an expert advisory panel in a bid to coordinate the drafting of ethics guidelines for human genome research. The goal is to bring some order to a growing number of regulatory proposals issued by government agencies by establishing a common ethical ground for all such research in Japan.

According to government officials, the panel is expected to draft model guidelines “as soon as possible”, and no later than December. Although these guidelines would not be binding, individual agencies would be expected to use them as a reference point.

The panel has been set up by the four agencies that fund most of Japan’s human genome research: the Science and Technology Agency (STA), the Ministry of International Trade and Industry, the Ministry of Health and Welfare (MHW) and the Education Ministry (Monbusho). Most panel members come from the agencies’ advisory committees.

The country’s Millennium Project — a five-year plan targeting areas of science and technology with high economic potential — has boosted funding for biomedical research, especially genomics. Following this, committees at various government agencies have drafted proposals for regulating the use of human material in research and protecting genetic information.

Advisory committees under the STA and MHW released guidelines for human genome research earlier this year. But so far regulation has been limited in scope and often applies only to research funded by a single agency or to activities within the Millennium Project.

An informal ‘genome constitution’ approved in late June by a subcommittee of the Council for Science and Technology, an advisory body close to the STA, has so far attracted little support from other agencies.

Critics in Tokyo argue that having the genomics guidelines drafted by an informal expert body without any legal status reflects fundamental problems in the Japanese approach to regulating biomedicine and biotechnology.

Flaws in the country’s regulatory approach were revealed earlier this year when its parliament rejected a bill drafted by the STA to prohibit human cloning. Critics argued that the content of the proposed law was inadequate and its wording scientifically contentious.

According to Shohei Yonemoto, a policy analyst at the Mitsubishi Kasei Institute of Life Sciences in Machida near Tokyo, the committees drafting research guidelines in Japan typically lack the “capability, autonomy and legitimacy” needed in a field of regulation that is exposed to public scrutiny.

Japanese industry has kept a low profile. As the international standardization of approval procedures for clinical trials enables Japanese pharmaceutical companies to use data from trials in Europe or the United States in applying for drug approval in Japan, companies such as Yamanouchi have decided to carry out clinical trials in areas such as pharmacogenomics abroad.

Various industry organizations, including the Japan Biotechnology Association and the Japanese Pharmaceutical Manufacturers Association, have issued broad statements on the ethics of human genome research. But companies have not moved into areas with an unclear regulatory regime. ■

The sky’s the limit as radio telescope array is approved

Steve Nadis, Boston

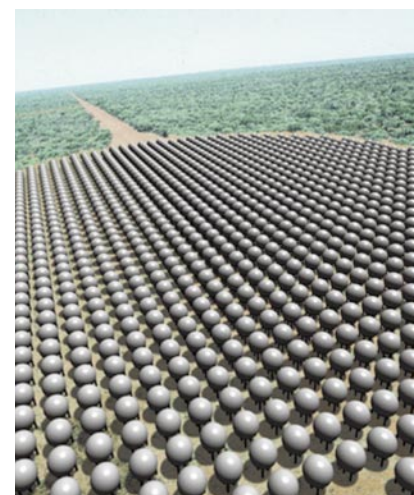
Small might be beautiful, but to astronomers bigger is often better. This is certainly true in radio astronomy, where international teams have now approved two giant arrays — the Atacama Large Millimeter Array (ALMA) and the Square Kilometer Array (SKA).

ALMA, scheduled to be built by the end of this decade on a high plateau in Chile’s Atacama Desert, will have 64 12-metre antennas tuned to millimetre wavelengths, making it the world’s largest ground-based telescope. But this will be dwarfed by SKA, a centimetre-wave radio array with a collecting area of a million square metres.

SKA, long a dream of radio astronomers, came a step closer to reality last week. At a meeting of the International Astronomical Union in Manchester, England, representatives from Australia, Canada, China, India, the Netherlands and the United States signed an agreement to bring SKA online by the middle of the next decade. The plan calls for choosing a design by 2005 and starting construction in 2010.

Why the continual push for bigger telescopes? “This project can be justified by the rule of thumb that says you need something ten times bigger to make major discoveries in astronomy,” says Barry Turner of the National Radio Astronomy Observatory in Charlottesville, Virginia. “Three times bigger can get you better data, but ten times bigger might lead to important new findings.”

Many challenges, including fund-raising, lie ahead. “There is no pile of



Square eyed? Artist’s impression of an array station, part of the Square Kilometer Array.

▶ money sitting on the table,” says University of Manchester astronomer Peter Wilkinson, an early champion of SKA. “What we have is a resolve to pool our resources and a consensus that [after ALMA] this is the next big project in radio astronomy.”

The ultimate location of the array is uncertain, with sites in Australia, Chile and China being considered. The success of the project will depend on whether researchers can find ways to reduce the cost by a factor of ten or more compared with conventional approaches, bringing the total to roughly a billion dollars.

There are several potential designs. Backers of the Allen Telescope Array, a one-hectare array named after the Microsoft billionaire Paul Allen who is giving it substantial funding (see *Nature* 406, 551; 2000), hope it will be a model for SKA. This telescope, which will link between 500 and 1,000 commercial radio dishes, is being developed by the SETI Institute and the University of California, Berkeley, for operation in 2005.

Australian researchers are investigating the use of ‘refracting’ telescopes called Luneburg lenses that would be deployed in 100 stations, each consisting of 400 lenses. The Canadian proposal is an antenna with a diameter of 200 metres that would reflect radio waves to a receiver suspended from a balloon 500 metres above the surface.

Chinese astronomers are contemplating a network of 30 radio antennas similar to the Arecibo telescope in Puerto Rico, a dish with a 300-metre diameter that sits in a natural depression. Researchers in the Netherlands favour a so-called phased-array approach, consisting of millions of small antennas embedded in flat panels.

Given the technical and logistical uncertainties, Turner, for one, is sceptical of the timetable set for SKA at the Manchester meeting. “For every step of the way, there are so many choices — the size of the dish, the target frequency range — all of which are subjects of sharp debate. They could get it all sorted out by 2005, but I wouldn’t count on it.”

But many in the field are confident that SKA will be worth the wait, even if it does not stick precisely to the schedule. For example, scientists are keen to spot irregularities in the distribution of the primordial hydrogen that gave rise to early galaxies.

And the telescope will also aid the search for extraterrestrial intelligence, enabling astronomers to examine more than a million Sun-like stars while listening for ‘leakage’ — random TV and radio signals — as well as messages sent intentionally. ■

NASA pins hopes on bigger, costlier mission to Mars

William Triplett, Washington

Is the era of ‘faster, better, cheaper’ space missions — NASA’s mantra for the past decade — over? Last week the US space agency announced a new mission to Mars that will cost almost \$600 million, roughly twice the combined cost of its previous two missions to the planet.

The announcement has led to speculation that there has been a fundamental shift in the agency’s thinking. It comes in the wake of an independent report, commissioned by NASA following recent failed missions, that questioned the ‘faster, better, cheaper’ policy.

Ed Weiler, NASA’s associate administrator for space science, the division in charge of the new mission, says that there has been a shift in attitudes, but not a big one. “It’s certainly an adjustment,” he says. “But it’s not a step back to the old days, when you did [a project such as] Viking and then waited ten years until you did anything else.”

Like the pair of Viking landers that NASA successfully sent to Mars in the 1970s, the new mission will dispatch a pair of rovers, each about the size of a golf cart, in separate launches in 2003. If one fails, there is still the other; if both succeed, the mission gathers more scientific information.

NASA is desperate to avoid a repeat of its most recent mission to Mars, last year’s loss of the Mars Polar Lander and Mars Climate Orbiter, which together cost \$300 million. The failures led to a review of NASA mission planning that concluded, as Weiler puts it, “that clearly ‘faster, better, cheaper’ had been pushed to its limits”.

The successful Mars Pathfinder mission,



Weiler: acknowledges a shift in attitudes.

which sent a single rover to the planet’s surface in 1997, cost about \$300 million. “We then took the bolder step of doing two missions for the price of one,” says Weiler. When these both failed, he adds, “it didn’t take a rocket scientist to figure out that maybe we’d pushed too far on costs and raised risks.”

But ‘fast, better, cheaper’ has not been completely ditched. Weiler points out that when adjusted for inflation today’s cost of the Viking missions, with landers but no rovers, would be \$2–3 billion. “So is the new Mars mission still cheaper? You bet.”

He also says that this is more than just a doubled repeat of Pathfinder. The mission’s goal is to search for signs of water, either present or past, which scientists think might be on Mars. But Pathfinder, Weiler notes, covered only 100 metres of territory in three months. The new rovers “cover about a hundred metres a day”.

Louis Friedman, executive director of the Planetary Society, agrees that the new mission signals an apparent shift in approach. But he is cautious about how significant the shift is. “The pendulum is definitely swinging back,” Friedman says. “But where it will end up, I’m not sure.”

He says that “all the space science news I’ve been hearing out of NASA is that a number of projects will be costing more now because of [the steps being taken] to ensure better reliability of space science missions. And it’s pushing costs up to the point that there’s a threat that other missions might have to be cancelled.”

There have already been warnings about possible threats to the Pluto–Kuiper Express (see *Nature* 406, 554; 2000). If too many such missions are cancelled, Friedman says, NASA may indeed find itself back in the days of Viking, with missions few and far between — less expensive, perhaps, but the same in principle.

But John Logsdon, director of the Space Policy Institute at George Washington University in Washington DC, doubts whether there will be even a small return to NASA’s old days. The new Mars mission, he says, is nothing more than “a mid-course correction based on hard experience”. The biggest implication of last week’s announcement, he argues, is that it demonstrates NASA’s continued commitment to exploring Mars. ■



One of a pair: NASA will send two rovers to Mars in 2003 to search for signs of water.

AFP/CORBIS

NASA

Science champion joins US election race

Colin Macilwain, Washington

For most Americans and foreign observers, Al Gore's selection of Senator Joseph Lieberman (Democrat, Connecticut) as his running mate in November's presidential election was noteworthy for the senator's religious affiliation: he is an orthodox Jew.

But for the scientific community, it means something else. Lieberman, a 58-year-old lawyer, is one of science's strongest supporters in the Congress. And if the Gore–Lieberman ticket triumphs, he is likely to play an active role in shaping US science policy.

Lieberman is one of perhaps half-a-dozen senators with a genuine interest in research issues. He helped to found a science and technology caucus in the Senate, and led congressional efforts to pass bipartisan legislation endorsing the idea of doubling the budgets of all of the major civilian research agencies over the next 10 or 12 years.

Lieberman introduced the first 'doubling bill' with Phil Gramm (Republican, Texas) in 1997, and later supported a more recent version sponsored by Bill Frist (Republican, Tennessee) and Jay Rockefeller (Democrat, West Virginia).

The Senate passed this measure unanimously last October, and the House of Representatives may consider it next month. Scientific societies see it as an important statement of support for agencies other than



Dream team? Both Lieberman (centre) and Gore (right) are strong advocates of science.

the National Institutes of Health, which is already assured of congressional backing.

Michael Lubell, head of public affairs at the American Physical Society, describes Lieberman as "the primary ingredient" in the doubling bill's progress, because of his ability to reach across party lines to build support. "He's got a very strong interest in science and technology, and complements Gore extremely well in that regard," he says.

The Connecticut senator is a founding member, with Frist, Rockefeller and Pete Domenici (Republican, New Mexico), of the science and technology caucus in the Senate, an informal group that meets about once a month and serves as a sounding-board for science and technology issues.

Jack Gibbons, a former science adviser to President Bill Clinton, says that Lieberman "appreciates the aesthetic as well as the economic value of research" and "has always worked hard for the research and development budget".

Lieberman had an early interest in arms control, writing a book on it in 1970. But, according to several sources, he developed an interest in technology policy largely through his involvement in small business and security issues, both of which reflected the economic needs of his home state.

Later, he became a champion for basic scientific research. He sits on the Senate Armed Services Committee and recently convened a conference between generals and scientists to discuss the future of warfare.

"He's been a long-time advocate of science and technology," says David Schutt, head of legislative affairs at the American Chemical Society, adding that Lieberman is involved "whenever there is a bill out there" to boost research.

No one expects science and technology to play much of a role in the presidential campaign. George W. Bush, the Republican candidate, has already promised to double military research and development if he is elected, and both parties will portray themselves as strong supporters of research.

But observers in Washington point out that the Gore–Lieberman ticket is an unusual combination: two lay politicians who take more than a passing interest in science.

Chemist tipped for top UK science post

David Dickson, London

One of Britain's leading physical chemists, well known in his field but less prominent in policy-making circles, is tipped to become the UK government's next chief scientific adviser and head of the Office of Science and Technology.

David King, master of Downing College, Cambridge, and head of the department of chemistry at the University of Cambridge since 1993, is said to have been picked for the post from a short list of three.

No formal offer has yet been made to King, who is currently on holiday. But time is pressing — the five-year contract of the present chief scientist, Sir Robert May, expires at the end of August, when he will start his new job as president of the Royal Society.

Born in South Africa, King studied at the University of Witwatersrand and at Rand



King: the next chief scientist?

University before coming to England. He was a lecturer at the University of East Anglia in Norwich before being appointed professor of physical chemistry at the University of Liverpool in 1974. He has been at Cambridge since 1988.

During his time at Cambridge, King, a specialist in the physics and chemistry of solid surfaces, has held several senior posts. He was a senior research fellow at the Royal Society of Chemistry, and has been an editor of *Chemical Physics Letters* since 1989. He has also held a number of advisory appointments and teaching commitments in Europe, and became a Fellow of the Royal Society in 1991.

Despite his low profile in government circles, he is no stranger to the political issues concerning university researchers, having been on the national executive of the Association of University Teachers from 1970 to 1978, and the union's president in 1976–77.

"Being science adviser sounds like his sort of job," says one chemist. "He is very good at running things, and has a reputation as a keen and efficient administrator." If King's appointment is confirmed, he is believed to be the first chemist to hold the post since it was created.

US dispute over definition of animal distress

Jessica Netting, Washington

Changes to animal welfare rules proposed by the US Department of Agriculture (USDA) have been met with concern from a group of US researchers. The scientists last week urged the agency to replace its current working definition of distress in laboratory animals, which they describe as “subjective” and “unenforceable”.

The researchers had taken part in a workshop held by the Federation of American Societies for Experimental Biology (FASEB) to set a policy agenda on animal welfare. The meeting followed a call from USDA in July for comments on proposed rules for classifying animal pain and distress in the laboratory.

USDA is seeking suggestions for an official definition of ‘distress’. It has also outlined alternative ways to measure pain and distress in animals. One idea is a three-tier classification of procedures — none/mild, moderate, severe — based on the level of pain experienced by the animal over the whole trial.

This categorization is similar to systems used in countries such as Canada and Switzerland. But many researchers would prefer a method, based on the current rules, that focuses on whether or not analgesic or anaesthetic drugs are used to alleviate pain

during a particular period of the trial.

Controversy over pain has been smouldering for years, says Ron DeHaven, deputy administrator for animal care with USDA’s Animal and Plant Health Inspection Service.

Both researchers and animal rights advocates point to areas of confusion or ambiguity in the present rules, stemming from a failure to define terms or clarify the categories to be reported. The word ‘distress’ appears as often as ‘pain’ in the Animal Welfare Act of 1966, but its exact meaning is never specified.

DeHaven says that this was more by oversight than intent. Unofficially, the department defines distress as “a state in which an animal cannot escape from or adapt to the internal or external stressors or conditions it experiences, resulting in negative effects on its well-being”.

Many participants in the FASEB workshop prefer another definition already adopted by the Institute for Laboratory Animal Research in Washington DC. This describes distress as “an aversive state in which the animal is unable to adapt completely to stressors and the resulting stress and shows maladaptive behaviour”. Maladaptive behaviour usually means self-mutilation, such as an animal chewing off the tip of its tail.

Animal advocate organizations say the def-



Painful question: what is distress for a lab mouse?

inition is too conservative. “By the time an animal starts to exhibit maladaptive behaviours, it may have been under stress for some time,” says Rick Bogle of In Defense of Animals.

But Kathryn Bayne, associate director of the Association for the Assessment and Accreditation of Laboratory Animal Care, says that the phrase “negative effects on its well-being” is too general to be enforceable. ■

► <http://www.aphis.usda.gov/ppd>

P. GONTIER/EURELIS/STL

Livermore laboratory kept fusion troubles quiet for a year

Washington The Lawrence Livermore National Laboratory in California knew of severe cost overruns on the National Ignition Facility in summer 1998, a full year before it told the government that the troubled laser fusion project was off track, says a report by the US General Accounting Office to be released this week.

In its long-awaited report to Congress, the office says that the facility will cost \$3.9 billion, including associated research projects. The Department of Energy, which is paying for the construction, estimates that the full cost will be \$3.3 billion. Originally, the department had planned to spend \$1.2 billion on construction and \$1 billion on research.

The Congress will decide whether to give the department the extra money that it needs when it returns from its summer recess early next month.

New head for French agricultural research body

Paris The French cabinet has picked Marion Guillou, currently director-general for food matters at the French ministry of agriculture

and fisheries, to succeed Paul Vialle as director-general of the National Institute of Agricultural Research (INRA).

An environmental engineer and physical chemist, Guillou has held several agricultural and research posts in the French government, including agricultural attaché at the French Embassy in the United Kingdom and director of research at the University of Nantes.

Sandage and Peebles share cosmology prize

London The first recipients of the Cosmology Prize of the Peter Gruber Foundation were announced this week. They are Allan Sandage, staff astronomer emeritus at the observatories in Pasadena, California, of the Carnegie Institution of Washington, and Phillip Peebles, Albert Einstein professor of science at Princeton University. The prize is the first award to be dedicated to cosmology, and each carries a cash award of US\$150,000.

Sandage was honoured for his observational contributions to such goals as estimating the true values of the Hubble constant and the age of the Universe. Peebles has more than three decades of experience in work to understand phenomena ranging from the creation of the lightest elements to the formation of galaxies and the cosmic distribution of matter and radiation. The

awards were announced during the general assembly of the International Astronomical Union in Manchester.

♦ <http://www.gruberawards.org>

Biologists get their hands on strongest magnet yet

London The most powerful magnet suitable for use in biological research has been developed and built by the UK company Oxford Instruments. The 900-megahertz magnet will be used for research into the three-dimensional structure of macromolecules such as DNA and proteins. It will be used in nuclear magnetic resonance systems supplied by Varian (see below).



OXFORD INSTRUMENTS

New fronts in an old war

There has been no new treatment for tuberculosis for three decades. But there is now the potential for a radical resurgence of drug development, says Declan Butler, if the political and industrial climate stays fair.

Next time you are in a busy street, glance around. If you live in North America or Western Europe, one in ten of those walking past you is probably infected with *Mycobacterium tuberculosis* — the bacterium that causes TB. And with infection rates in developing countries and Eastern Europe soaring, the global average is close to one in three.

Thankfully, the virulence of *M. tuberculosis* does not match its prevalence: less than a tenth of those infected will develop the full-blown disease. In the rest, the bacterium seems to lie dormant. But the number of cases where this latent TB becomes activated is increasing, mainly because the rise in HIV infection is weakening the immune systems of those also carrying *M. tuberculosis*. At the same time, many strains of the bacterium are developing resistance to the drugs used to treat the disease. Together, these trends have led the World Health Organization (WHO) to declare a global health emergency (see 'The rise and rise of tuberculosis', opposite).

"We need new drugs," concludes the draft of a scientific blueprint being prepared by the Global Alliance for TB Drug Development, a consortium of scientists, donor agencies and pharmaceutical companies set up earlier this year to provide a new impetus for the field¹. The document will be released in the autumn.

The good news is that recent scientific developments, including the sequencing of the *M. tuberculosis* genome², mean there is a chance of making significant progress. "This is the most exciting time for TB drug development in over two decades," says Rick O'Brien, chief of research in the Division of

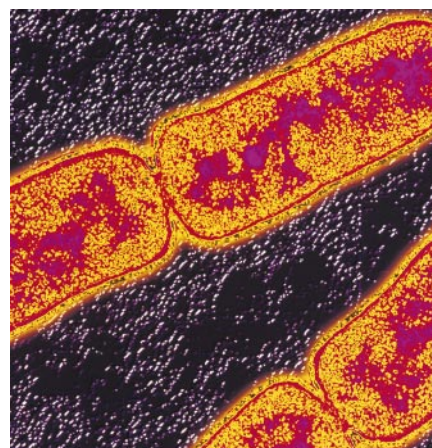
TB Elimination at the Centers for Disease Control and Prevention in Atlanta, Georgia, and rapporteur for the alliance's scientific blueprint. "At long last technical advances and increasing interest in the global TB problem are coming together."

Traditional treatment

TB is usually treated with a cocktail of four drugs — ethambutol, isoniazid, pyrazinamide and rifampicin. These drugs are ineffective against latent TB, but work well when *M. tuberculosis* is actively dividing. Two months of treatment can reduce infection in sick patients to the point where they can no longer transmit the disease. But in the face of this pharmaceutical assault, the remaining bacteria slow their cell division and enter a persistent state that resembles latent infection. It can take another six months of treatment to kill these remaining bacteria.

For this reason, the WHO devised a scheme called DOTS, or Directly Observed Treatment, Short-course. In DOTS, doctors actively diagnose TB cases and subsequently ensure that patients swallow each and every pill — which may also include the broad-spectrum antibiotic streptomycin — over six to eight months. But DOTS requires trained medical staff, infrastructure and money that lie beyond the resources of many developing countries. In practice, a quarter of TB cases are not diagnosed before they become infectious, and many people given the treatment do not stay the course.

The Global Alliance for TB Drug Development says its priority is to develop a drug that will speed up elimination of the slow-growing bacteria that initially survive the drug cocktail. Ideally, the alliance wants



The enemy within: *Mycobacterium tuberculosis*.

to reduce the total time needed for treatment to two months or less, says Ariel Pablos-Mendez, associate director for health equity at the Rockefeller Foundation in New York. This would be a revolution in TB control, he adds.

Persistent problem

Given the similarities between the persistent bacteria and those involved in latent infections, such a drug might also prove effective in eradicating *M. tuberculosis* from asymptomatic TB-positive people. Many researchers agree that studying slow-growing *M. tuberculosis* is a top priority. "If you could find a drug that would kill these persistent organisms that would be an enormous step forward," says Denis Mitchison, now retired, but continuing to carry out research at St George's Hospital in London, where he worked on the clinical development of today's TB drugs.



Kill or cure? More than one in three people in the developing world is infected with TB and unless new drugs are found to fight the disease, many will die.

ALFRED PASTEK/SPL

LIBA TAYLOR/PANOS PICTURES

Alternatively, it might be possible to force the bacteria to start dividing rapidly again, so that other drugs could do their work. "We could prevent latency, or even stimulate persistent or dormant bacilli to grow and be better able to kill them," says William Bishai of Johns Hopkins University in Baltimore, Maryland.

New families of drugs that can target dividing *M. tuberculosis* are also urgently needed to sidestep growing bacterial resistance to isoniazid and rifampicin. Multiple-drug-resistant TB accounts for around 2% of all cases worldwide, and this figure is increasing. The proportion of cases that are multidrug resistant has reached 14.1% in Estonia, and more than 25% in Russia's prisons.

The problem confronting drug developers is that decades of neglect in TB research — partly fostered by the misconception that existing drugs had beaten the disease — have left severe gaps in our knowledge of the biology of *M. tuberculosis*. "We know singularly little about pathogenesis," says Mitchison.

Bright prospects

But morale among TB researchers is improving. There is a raft of new scientific opportunities and *M. tuberculosis* has a host of unique features that, in principle, should make it a relatively easy target for drugs. In addition, there is renewed political and public interest worldwide in funding TB research. For example, at last month's summit in Okinawa, Japan, the G8 economic powers pledged to halve the number of TB deaths and the prevalence of disease by 2010.

Much of the impetus for research has come from the availability of the *M. tuberculosis* genome. Functions have been identified for almost half of the bacterium's genes², and genomic approaches are now being applied to drug discovery. At the Affymax Research Institute, a biotech company in Santa Clara, California, Mike Wilson is using microarrays of the bacterium's 4,000 known genes to investigate the effects of TB drugs on gene expression³. By adding 'complementary' DNA, which is made from the bacterium's messenger RNA and binds to the corresponding genes on the microarray, he can see which genes are active before and during treatment.

Wilson is now working with Clifton Barry's group at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, to speed up screening of candidate drugs by profiling gene expression patterns specific to particular biochemical pathways — starting with the production of *M. tuberculosis*'s characteristic waxy cell wall. "These are complex structures that require many biosynthetic steps, so we might expect to find a variety of potentially useful targets if we could identify the set of genes responsible," says Wilson. Initially, they are concentrating on finding new families of drugs that have similar effects on gene expression to isoniazid, which

is known to disrupt cell-wall formation.

But latent TB remains a mystery. 'Latent' is a clinical definition, meaning simply that the bacteria are present but there is no active disease. Biologists do not know whether the bacteria are lying dormant or are actively replicating but held in check by the immune system. "I prefer the term 'persistent and chronic TB infection,'" says David Russell of Cornell University in Ithaca, New York. "The terms dormancy and latency imbue the bacterium with properties that have not really been demonstrated."

Cell in-mates

One hypothesis is that the bacteria hide inside immune cells known as macrophages. These are large cells that ingest foreign microorganisms and cellular debris in diseased tissues. Unlike other bacteria, *M. tuberculosis* can survive inside these cells, where it initially grows and divides. Over time, the immune system walls off infected macrophages into structures called granulomas, and the bacteria slow down their division. There they can lie for many years before suddenly resuming their rapid multiplication and bursting out of the granuloma to cause active disease.

Inside a granuloma, *M. tuberculosis* has to adapt to declining levels of oxygen. To investigate this, Lawrence Wayne of the Long Beach Veterans Affairs Medical Center in California has devised an experimental model in which oxygen levels in *M. tuberculosis* cultures are gradually reduced so that the bacteria survive but do not divide⁴.



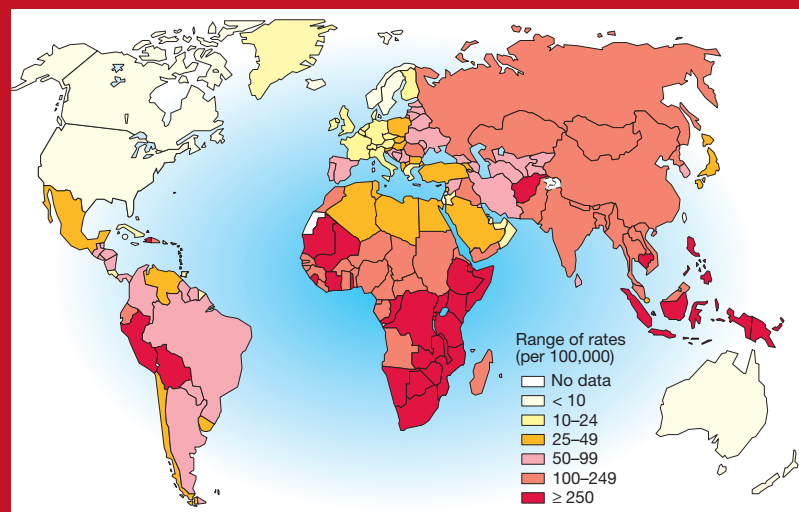
Other researchers are studying infected mice which develop a chronic infection similar to latent human TB. But mouse granulomas are different from those seen in humans — they do not form structures that can eventually rupture into a bronchus, allowing virulent bacteria to be coughed out. Guinea pigs⁵ and rabbits⁶ provide better models, but it is more difficult to obtain large inbred populations and, unlike mice, these animals can spread the disease to humans. As a result, they can only be studied in expensive isolation facilities.

Drugs on target

Although all of the models have shortcomings, they have helped to identify many new drug targets for persistent and chronic forms of *M. tuberculosis* over the past year or so. Among the most exciting is the discovery, announced in this issue of *Nature*⁷ by a team led by Russell, that an enzyme called isocitrate lyase seems to be crucial to the survival of *M. tuberculosis* in granulomas. This enzyme allows the bacteria to get energy and build carbohydrates from fatty acids in the absence of oxygen.

Russell and his colleagues, who this month also reported the crystalline struc-

The rise and rise of tuberculosis



Cases of active TB per 100,000 population for 1997 (source: ref. 13).

- 8 million new cases of active TB annually, including 3.5 million cases of infectious pulmonary disease
- 2 million of these — 5,000 per day — will die
- 16 million existing cases of active TB
- Without new control measures, one billion new cases of TB are predicted between now and 2020, of whom 200 million will develop active disease, and 35 million will die. Annual incidence of new cases will grow to 11 million

ture of isocitrate lyase⁸, are now working to develop inhibitors of the enzyme. The UK-based drug company Glaxo Wellcome is also taking an interest. "We have already begun to look for inhibitors of the isocitrate lyase enzyme in our library of compounds, using the structure of the protein to guide the search," says Ken Duncan, manager of Glaxo Wellcome's Action TB programme.

Another clue to the genes involved in persistent infection came earlier this year from Stanley Falkow's laboratory at Stanford University in California. Falkow and his colleagues studied *M. marinum*, which causes latent TB in frogs and is the closest relative of *M. tuberculosis*. They identified gene promoters that were only switched on when the bacteria were encapsulated in frog granulomas. Two of these were from genes in the PE-PGRS family, and when the researchers examined the sequence of the genomic region containing these genes they found a third family member⁹. PE-PGRS genes code for unusual proteins that are rich in the amino acid glycine but have unknown function. They are unique to the genus *Mycobacterium* and account for more than 10% of the coding regions of its genome².

Falkow and his colleagues created strains of *M. marinum* in which each of the three genes was knocked out individually. Two of these knockout strains could not divide in macrophages, and one showed greatly reduced persistence in frog granulomas. The group now intends to knock out all three genes in unison to see whether this will reduce the bacterium's persistence further. The researchers hope that similar genes in *M. tuberculosis* will play the same role in human TB, and therefore make useful drug targets. "We know that these genes have a role in virulence," says Lalita Ramakrishnan, a member of the team.

Hot on the heels of Falkow's paper came the news, from a group led by Kendall Stover

at PathoGenesis, a company in Seattle, that molecules called nitroimidazopyrans seem to kill *M. tuberculosis*, including persistent, slow-dividing bacteria¹⁰. This discovery stemmed from observations that related molecules called bicyclic nitroimidazofurans, originally developed to improve the effectiveness of cancer radiotherapy, were active against *M. tuberculosis*. Those drugs were too mutagenic to use against TB, however.

Survival strategies

Meanwhile, researchers led by Jean Pieters at the Basel Institute for Immunology in Switzerland have been looking at how *M. tuberculosis* bacteria living inside macrophages avoid attack by lysosomes — cell structures containing enzymes that break down bacteria and debris ingested by the macrophages. Last year, they discovered that a macrophage protein called tryptophane aspartate-containing coat protein, or TACO, seems to play a key role¹¹. When the bacteria are ingested by a macrophage they end up surrounded by a host-cell membrane, forming a structure called a phagosome. The researchers found that TACO surrounds these phagosomes, preventing transport of the bacteria to the lysosomes.

Pieters wanted to discover how to release TACO from the phagosomes. After testing several chemicals, he found that digitonin, a solubilizing agent, completely released the protein¹². It occurred to Pieters and his colleague John Gatfield that digitonin also binds to cholesterol. "As we knew TACO was responsible for mycobacterial survival within phagosomes, we reasoned that if we get rid of the cholesterol, TACO cannot be recruited to the phagosome and the mycobacteria will be transferred to lysosomes and killed," says Pieters.

When the researchers tested this idea in cultures of macrophages, they found that reducing cholesterol had a second, even more dramatic effect. "To our surprise we found that the bacteria were no longer taken up by the macrophages," says Pieters. "Cholesterol thus has dual roles: getting the bacteria to be taken up, and then conferring protection from lysosomes."

It would not be feasible to reduce cholesterol levels as a therapy for persistent TB. But Pieters believes it might be possible to identify cholesterol-binding sites on *M. tuberculosis*, block them with drugs and so prevent the bacteria entering macrophages. He and his colleagues are now planning to look at mutant bacteria that cannot get into macrophages to work out which genes are involved.

TB researchers are optimistic that continued research will generate many more potential drug targets. But the route from a target to a drug is long and expensive, and the big unknown is whether drug companies will invest the multimillions of dollars needed for



This is the most exciting time for tuberculosis drug development in over two decades **Rick O'Brien.**

drug development. Realizing this, the Global Alliance for TB Drug Development is preparing a report on 'pharmaco-economics' alongside its scientific blueprint. This will assess in detail the size of the TB drug market, current investment and gaps in corporate research and development. It plans to use this report as the basis for a business plan for a joint drug-discovery programme between the public and private sectors.

The alliance's stated goal is to register one new drug for chronic and persistent TB by 2007, and a second by 2012. With the difference between success and failure being measured in millions of human lives, the stakes could not be much higher.

Declan Butler is Nature's European correspondent.

1. Butler, D. *Nature* **403**, 692 (2000).
2. Coles, S. T. et al. *Nature* **393**, 537–544 (1998).
3. Wilson, M. et al. *Proc. Natl Acad. Sci. USA* **96**, 12833–12838 (1999).
4. Wayne, L. G. & Haynes, L. G. *Infect. Immun.* **64**, 2062–2069 (1996).
5. O'Grady, F. & Riley, R. L. *Adv. Tuberc. Res.* **12**, 150–190 (1963).
6. Dannenberg, A. M. in *Tuberculosis: Pathogenesis, Protection and Control* (ed. Bloom, B. R.) 149–156 (American Society for Microbiology, Washington DC, 1994).
7. McKinney, J. D. et al. *Nature* **406**, 735–738 (2000).
8. Sharma, V. et al. *Nature Struct. Biol.* **7**, 663–668 (2000).
9. Ramakrishnan, L., Federspiel, N. A. & Falkow, S. *Science* **288**, 1436–1439 (2000).
10. Stover, C. K. et al. *Nature* **405**, 962–966 (2000).
11. Ferrari, G., Naito, N. H., Langen, J. & Pieters, J. *Cell* **97**, 435–447 (1999).
12. Gatfield, J. & Pieters, J. *Science* **288**, 1647–1651 (2000).
13. Dye, C., Scheele, S., Dolin, P., Pathania, V. & Ravigione, M. C. *J. Am. Med. Assoc.* **282**, 677–686 (1999).



Shadow of death: TB lesion in the lung.

Though HIV is the main cause of AIDS, other factors play a role

Sir—Although I entirely endorse the stand taken by the authors of the Durban Declaration¹, I think that HIV is the principal, rather than the sole, initiating factor of the current world pandemic of AIDS. I believe this for the following reasons.

First, from my experience of secondary or acquired immune deficiency in treated lymphoma patients who are HIV negative and in patients who have chronic graft-versus-host disease after a transplant, HIV is not the only mechanism for adults developing the clinical syndrome. Second, external influences such as immune stimuli can modify progression, as demonstrated when haemophiliacs given recombinant factor VIII have a slower loss of CD4 lymphocytes than those who receive immunogenic serum protein containing factor VIII concentrates². Third, poor nutritional status identified on the basis of low vitamin A levels is associated with increased viral shedding³. This augments HIV perinatal transmission rate in breastfeeding African women⁴ and non-breastfeeding US women⁵. Fourth, general genital hygiene⁶, underage sexual activity and circumcision⁷ all affect the risk of acquiring HIV infection.

It is appalling that, as Dorritie states in Correspondence⁸, there may be a right-wing political agenda behind the campaign claiming AIDS is not caused by HIV. Nevertheless, other social factors are involved in the African aspects of this epidemic, such as Johannesburg's high incidence of rape, the myth that raping a virgin can cure AIDS, and the high level of non-HIV sexually transmitted diseases in sub-Saharan Africa⁹.

As a consequence, there are benefits to be gained from President Mbeki's call for scientists to address the nature of AIDS and its links to HIV in an African setting. Such an investigation could well conclude that investments such as educating young men in the basics of sexual hygiene, providing nutritional supplements for pregnant women and treating mastitis¹⁰ might produce greater gains than spending money on antenatal treatment with antiretroviral drugs.

R. T. D. Oliver

Department of Medical Oncology, Saint Bartholomew's Hospital, London EC1A 7BE, UK

1. *Nature* **406**, 15–16 (2000).
2. Hoot, K. & Canty, D. *Haemophilia* **4**, 704–713 (1998).
3. Mostad, S. B. *et al.* *Lancet* **350**, 922–927 (1997).
4. Semba, R. D. *et al.* *Lancet* **343**, 1593–1597 (1994).
5. Greenberg, B. L. *et al.* *AIDS* **11**, 325–333 (1997).
6. Beral, V. *et al.* *Lancet* **339**, 632–635 (1992).
7. Halperin, D. T. & Bailey, R. C. *Lancet* **354**, 1813–1815 (1999).
8. Dorritie, D. *Nature* **406**, 13 (2000).
9. Gerbase, A. C. *et al.* *Lancet* **351**, 2–4 (1998).
10. Semba, R. D. *et al.* *J. Infect. Dis.* **180**, 93–98 (1999).

Responsible aquaculture can aid food problems

Sir—If agriculture were to arise today, knowing what we do about its impacts on the natural environment, biodiversity and poverty, it would probably encounter the same mixed reception as aquaculture. R. L. Naylor *et al.*¹ give an excellent overview of the serious challenges facing aquaculture. As they show, in some cases it actually reduces fish supplies for human consumption.

Naylor *et al.* list four main goals for the industry: more farming of fish at low trophic levels (low on the food chain); reduction of fish meal and fish oil inputs in feed; development of integrated farming systems; and promotion of environmentally sound practices.

After 23 years of aquaculture research in Asia, Africa and the Pacific, the International Centre for Living Aquatic Resources Management (ICLARM) strongly supports these goals and adds a fifth: access for poor consumers and small-scale producers. Leasing and community organization can allow even landless people to become fish farmers and conduct profitable micro-enterprises if aid is well-targeted.

Non-government agencies are carrying out many such projects in countries such as Bangladesh, where fish can be up to 80% of animal protein consumed. They farm various species of carp (most importantly *Catla catla*, *Labeo rohita* and *Cirrhinus mrigala*), catfish (*Clarius* spp.) and tilapia (*Oreochromis niloticus*), which require low input levels. In the Pacific, we are showing the potential for farming giant clams (*Tridacna* spp.), blacklip pearl oysters (*Pinctada margaritifera*), hard and soft corals and tropical marine aquarium fish, and for restocking and enhancing the sea cucumber *Holothuria scabra*.

Small island developing states rely on marine resources for their income. French Polynesia and the Cook Islands, for example, use environmentally friendly aquaculture to rear blacklip pearl oysters, which are low on the food chain and—producing prized black pearls—are second only to tourism for bringing in foreign exchange. The value of the black pearl harvest in French Polynesia in 1999 was US\$200 million. Demand for cultured giant clams, and hard and soft corals, is increasing, and these can be grown on coral reefs with apparently no adverse impact. All these species and systems have been selected either for accessibility to poorer people, for environmental sustainability, or to supply products that are used locally.

Rising prices mean that fish must be cheap to produce, such as carp and tilapia, or poorer consumers will not benefit. M. Dey *et al.*² show that adopting the improved strain of Nile tilapia will reduce tilapia price by 5–16% in Bangladesh, China, the Philippines, Thailand and Viet-

nam. Poorer consumers will benefit most in all these countries except the Philippines.

Aquaculture can also restock high-value coral-reef species that have been severely depleted by over-fishing. In the small islands of the Pacific, research has paved the way for inshore marine resources to provide substantial income for coastal communities. Methods have been developed for culturing sea cucumbers, giant clams, trochus and green snail, and the emphasis of research is now on how best to release juveniles into the wild.

Release of cultured juveniles can also overcome the widespread problem of too few young fish returning to settle in the fishery area. Such practices have been used for 90 species in Japan over the past 30 years, particularly successfully in the case of the scallop.

Aquaculture and stock enhancement are still in their infancy, and some mistakes have been made in their development. New directions should follow the guidance outlined by Naylor *et al.*; they should also heed the questions of access by poorer societies and proceed wisely with stock enhancement under suitably safeguarded conditions.

M. J. Williams, J. D. Bell, M. V. Gupta, M. Dey, M. Ahmed, M. Prein, S. Child, P. R. Gardiner, R. Brummett, D. Jamu
ICLARM, World Fish Centre, PO Box 500 GPO, 10670 Penang, Malaysia

1. Naylor, R. L. *et al.* *Nature* **405**, 1017–1024 (2000).

2. Dey, M. M. *et al.* *Aquacult. Econ. Mgmt* (in the press).

Will we ever know what the Chinese knew?

Sir—I was intrigued by recent letters about the Chinese medical treatise *Su Wên* and its apparent account of blood circulation. Pioreschi's argument (*Nature* **405**, 993; 2000), based on uncertainty about the exact meaning of certain words, is insufficient to disqualify the ancient Chinese as the discoverers of blood circulation. One's conclusion can be biased by the choice of translator as well as the translator's choice of words: an inevitable problem since this argument relies on excerpts of translated materials.

In order to make a fair judgement, one would have to consult the ancient Chinese text directly and treat the technical terms in accordance with standard definitions, as one would do with any other technical literature.

Until ancient Chinese knowledge of blood circulation is proved or disproved in an appropriate manner, it is safe just to say that Harvey made the suggestion that elicited the documented scientific experiments proving how the blood circulates.

Ichikawa Shinji

FKAMPO (Traditional Japanese Medicine journal) 540-29 Ikeno, Ikeda-cho, Ibi-gun, 503-2424 Gifu, Japan

Rights and wrongs

A call for apes to be given legal personhood, and to find their place in research.

Rattling the Cage: Towards Legal Rights for Animals

by Steven M. Wise

Profile Books/Perseus: 2000, 362 pp.

£12.99/\$37.95

Animals in Research: For And Against

by Lesley Grayson

British Library: 2000, 318 pp, £35, \$69.95

Kenan Malik

Are animals persons or things? That's the question at the heart of Steven Wise's controversial new book. Legally, Wise points out, only humans are 'persons' — animals are merely things. It is this distinction that Wise challenges. He demands "legal personhood for chimpanzees and bonobos" so that they can be recognized as "a potential bearer of rights". *Rattling the Cage* is a lucidly written, passionately argued case for animal rights. What makes it different is that Wise — a practising attorney who specializes in representing animals — builds a legal, as opposed to simply a scientific or philosophical case, for rethinking the relationship between humans and animals.

He begins by excavating the historical and philosophical roots of laws relating to animal rights. The Roman belief that 'all law is established for men's sake' remains central to the modern legal system. But this notion, Wise believes, has given rise to conflict between 'natural law' and the 'law of nations'. Natural law he defines as a universal law rooted in the nature of things, the law of nations as specific laws that humans make, and which can vary according to time and place. Slavery and racism, Wise suggests, were 'abhorrent' to natural law but acceptable under the law of many nations. Similarly, the denial of personhood to animals is a product of human law but contrary to natural law. The fundamental question we have to answer, according to Wise, is whether "things or beings or ideas [are] valuable because we value them or because they are inherently valuable".

This argument is surely wrong. Historically, slavery and racism were challenged not because they were incompatible with natural law but when they were seen as contrary to human morality. Indeed, the idea that something is inherently valuable lay at the heart of racial thinking: the belief that certain people were by nature more valuable than, or superior to, others. As the great Enlightenment philosopher the Marquis de Condorcet put it, racism "makes nature an accomplice in the crime of political inequality".

To demonstrate the inherent value of chimpanzees and bonobos, Wise provides an



Childlike? Some believe that an ape may have the cognitive skills of a three-year-old child.

overview of recent research on the mental capacities of the great apes. Apes, he argues, possess the same kinds of complex cognitive abilities as humans; they have minds and are able to read others' minds, possess consciousness and self-consciousness, and are capable of using language. Most of the material is familiar — the work of ethologists and primatologists such as Donald Griffin, David Premack, Frans de Waal, Michael Tomasello, Roger Fouts and Sue Savage-Rumbaugh. Wise, however, skates over the controversial nature of much of this work — especially the interpretation of the data about apes' linguistic and mind-reading abilities — and peremptorily dismisses any critics. Daniel Dennett's arguments are "bizarre", while those of Daniel Povinelli are "extraordinary for any ape researcher".

At the heart of Wise's approach is the argument from analogy. We can assume that apes are conscious, intentional beings, he argues, for the same reason we assume that humans are so. We do not have access to other people's heads, but most of us accept that other human beings have minds, beliefs and desires as we do, because they behave much as we do. But if we can make this assumption with humans, why not with animals? Wise adopts Frans de Waal's axiom — that "strong arguments would have to be furnished before we would accept that similar behaviours in related species are differently motivated" — to argue that the great apes have minds similar to those of human children.

It was the philosopher Ludwig Wittgen-

stein who showed why this argument is muddle-headed. Human minds could not be truly private, he pointed out, because if they were we would not be able to communicate the contents of our minds to anyone else. My inner feelings mean something to me, in some part at least, insofar as they mean something to others. I can make sense of my self only insofar as I live in, and relate to, a community of thinking, feeling, talking beings. Far from inferring other humans' experiences from our own, we can only truly know what goes on inside our own heads by relating to other humans. It is only because we live, not as individuals, but within a social community and, moreover, within a community bound together by language, that we can make sense of our own inner thoughts and feelings. No animal possesses either language or a social network like ours. Therefore, it is simply not valid to assume that they have inner experiences as we do.

Suppose it were true, however, that apes did have the kinds of minds Wise imagines they do — that they are conscious, intentional beings with the cognitive skills of a three-year-old child. Would this be sufficient to allow them to possess rights? Wise clearly thinks so. Just as we give rights to three-year-old children, so we should give rights to equivalent apes.

This, however, is to misunderstand the character of rights. Rights do not exist in the abstract, but are created by beings capable of possessing and asserting such rights. To have a right means also to be responsible for one's

actions, to be recognized as a *moral* being. Adult humans, unlike children or apes, live within a web of reciprocal rights and obligations created by our capacity for rational dialogue. We can distinguish between right and wrong, accept responsibility and apportion blame. Neither young children nor apes can do this, and hence cannot bear rights. In truth, children are not accorded rights but *protections*. A right requires us to make our own decisions. A protection requires us to make decisions on behalf of others. Indeed, many rights, such as the right to vote, are denied to children, precisely because they cannot make rational decisions.

Why does the law provide protection for children — and for the mentally disabled, who also may not be rational and autonomous — but not for apes? Because children normally grow up to be full members of the moral community. Indeed, it is only because we treat them as potential moral agents that they grow up to be so. As for mentally disabled people, we provide them protections because they once possessed the potential to be a moral being. Children and the mentally handicapped are of the same kind as adult, autonomous humans: the kind whose normal instance is a moral being. Apes are not.

Wise is right that there is a clear demarcation in law between 'persons' who possess rights and non-persons (or things) who don't. But his desire to break down this distinction confuses the categories of moral and non-moral beings. Apes, like many animals, are individuals, each with a distinguishable character. But humans are individuals in an entirely different sense: we are self-created beings, who realize ourselves only through our relations with other such beings. This is at the heart of what we mean by personhood, and why animals may be individuals but can never be persons.

The real impact of the campaign for rights for apes is to diminish rights for humans. The current controversy over increased applications for animal research in the wake of the sequencing of the human genome is a good case in point. The government is worried about the political impact of an increase in new licences in a pre-election year. The consequence will be that disease will remain untreatable for longer and people will continue to suffer from the effects of potentially avoidable ill-health and disability.

Whereas *Rattling the Cage* is a work of advocacy, and necessarily one-sided, *Animals in Research* is a bibliography of both sides of the debate. Compiled by Lesley Grayson for the British Library, the bibliography is both comprehensive and highly readable. Grayson covers a wide field, including debates about animal consciousness, the value of animal research and the issue of the law and of regulation, as well as newer issues such as the GM revolution and

transgenic animal welfare. The aim of the book, she writes, is to create a 'dialogue' between the two sides in the debate. Both the aim and the book itself are invaluable. ■

Kenan Malik is a freelance writer living in London, UK (e-mail: kenanmalik@msn.com).

Any old bones?

Skeletons in Our Closet: Revealing the Past Through Bioarchaeology

by Clark Spencer Larsen

Princeton University Press: 2000. 248 pp.

\$35, £22

Christopher Wills

The native peoples of North America have survived two traumatic events during the past two millennia. One we all know about: the arrival of Europeans with their freight of diseases and their ability to destroy whole ways of life almost overnight. The other, earlier event is less obvious: the transition from a foraging way of life to one based on grains, which happened more recently in North America than the agricultural revolutions in other parts of the world. Both events, as Clark Spencer Larsen demonstrates in this fascinating and well-written book, left many traces on the skeletons of the peoples who lived through them.

Larsen and his many collaborators have examined burial sites spanning these two traumatic periods. Burials in Stillwater Marsh, Nevada, were recently uncovered by flooding in the region and cover the time from 1300 BC to AD 1300. Isotopic analysis by Margaret Schoeninger shows that the people living there ate plants gathered from the marsh and animals that they hunted. They were strong and healthy, with few signs of

vitamin deficiency and virtually no tooth decay. Nonetheless, their teeth show signs of periodic famines.

The story was very different with pre-Columbian people living on St Catherine's Island, Georgia. Skeletons dated from 400 BC to AD 1450 show a transition to ever-higher proportions of corn in the diet, accompanied by a huge increase in tooth decay and changes in the bones that indicate a more sedentary life. Jaws became smaller with the introduction of corn mush into the diet, and teeth were more crowded. The teeth were slightly reduced in size during this period. Famine was less common than at Stillwater, but in spite of their more reliable food supply there is little evidence that these people lived any longer than those in Nevada.

The second great trauma had even more dramatic effects. Larsen found that burials from Spanish Florida spanning the arrival of Europeans show the impact of introduced diseases. The greater number of Retzius (growth) lines in the teeth shows that severe childhood illness rocketed. Signs of infection in the long bones, virtually unknown before the arrival of the Europeans, are found in 60% of the burials from AD 1600. Unrelenting toil in the fields, as the native people were effectively turned into slaves by their European masters, left marks in the shapes of their bones. The strength of the bones, particularly in women, was reduced.

The lives of all these people, both before and after the traumatic transitions that they went through, were nasty, brutish and short. Larsen shows, however, that they were nasty, brutish and short in very different ways. Skeletal changes over a span of centuries or even decades were pronounced. We think of the agricultural revolution as an unqualified benefit to our species, but Larsen demonstrates that it also had substantial negative



Boning down: the North American agricultural revolution resulted in pronounced skeletal changes.

KEVIN FLEMING/CORBIS

aspects. Overall health was reduced by both the introduction of agriculture and the arrival of the Europeans.

Larsen is on shakier ground when he suggests that these phenotypic changes resulting from changes in lifestyle have been an important part of the evolution of our species. Skeletons can alter dramatically from one generation to the next depending on use and disuse, but these changes can be purely phenotypic, with no effect on the genes of their owners. Did the gene pools of the native Americans change during this period? And were any changes the result of selection for survival under new circumstances, or simply the result of tribal migrations? Mitochondrial DNA analysis shows that the Stillwater people were a mixture of different tribal groups, and a change in that mix could have taken place over time even in the absence of natural selection.

The skeletons are silent on the subject of true evolutionary change, but DNA from them may provide us with some answers in the near future. We will have a far better chance of detecting any genetic changes that resulted from the North American agricultural revolution and the arrival of Europeans than from similar but earlier traumatic events in Europe, Asia and Mesoamerica, because they are far closer to us in time. There seems no doubt that, as our expertise grows, the study of native American skeletons that Larsen has pioneered so effectively will tell us even more about our evolutionary history. This is because it seems unlikely that such huge events as the agricultural revolution and invasion by alien cultures have been squeezed into such a short span of time at any other point in our five-million-year history. Even the replacement of Neanderthals by the Cro-Magnons and other modern Europeans took at least 10,000 years. My guess is that we will find that true evolutionary changes took place as a result of these traumas, and that the changes were pronounced. ■

Christopher Wills is in the Department of Biology, University of California, San Diego, La Jolla, California 92093-0116, USA.

From donkeys and cows to whales

Marine Mammals: Evolutionary Biology

by Annalisa Berta & James L. Sumich
Academic/Harcourt: 1999. 494 pp.
\$59.95/£39.95

Axel Meyer

The study of marine mammals is a highly fragmented and multidisciplinary endeavour, often conducted in an intensive, but seldom eclectic, manner. The discipline is unit-



Dancing in the blue

Dolphins (National Geographic, \$35, £20) by Tim Cahill follows the lives and careers of three scientists involved in dolphin research. Underwater photographs, such as the one above

showing a group of spotted dolphins, sit beside pictures of the scientists at work, and play, on land. The book documents their motivation, and their achievements.

ed solely by its objects of study, the 100 or so species of mammal that are considered to belong to the marine mammal club. Species from three mammalian orders make up the list: those from the Carnivora include the pinnipeds (seals, sea lions and walruses), sea otters and polar bears; those from the Cetacea include whales, dolphins and porpoises; and those from the Sirenia the dugongs and manatees.

These mammals spend most, if not all, of their lives in the sea. Interestingly, during evolution, the marine ancestors of dolphins in South America, India and China repeatedly and independently recolonized freshwater habitats. Sadly, and ironically, the study of cetaceans began in earnest only in the first part of last century, just as it was beginning to be recognized that they were vanishing from our oceans and sustainable catch quotas had to be established.

Various marine adaptations permitted warm-blooded, air-breathing mammals to leave the safe confines of land and return to the watery habitats of their distant ancestors. These include insulation (blubber) and circulatory adaptations such as countercurrent

heat-exchange systems to deal with the extra heat loss in cold water. The animals' eyes, nose, ears and limbs also became highly modified, sporting sensory adaptations and key evolutionary innovations such as echolocation, baleen to allow filter-feeding, collapsible lungs that allow deep and prolonged diving, a tail and flippers. In the cetacean lineage, the evolution of many of these adaptations for marine life was apparently extremely rapid — much faster than in the closely related artiodactyls (which include cows, pigs and giraffes) and perissodactyls (donkeys, rhinos and horses).

The study of marine mammals, and particularly that of dolphins and whales, shares with primatology the cachet (or stigma, depending on your point of view) of intense public interest. Romantic notions of 'talking with dolphins', saving whales and emulating the French oceanographer Jacques Cousteau have eager students flocking to this field. But reliable data collection and controlled experiments on marine mammals can be extremely difficult, if not impossible. The great rarity of some species, frequently as a result of whaling, and the animals' often enormous

body size prevent many experimental approaches. As a result, a lot of problems, particularly in behavioural ecology, are still unanswered. This remains the case in spite of advances in microelectronics such as satellite telemetry, and molecular approaches using the polymerase chain reaction and DNA fingerprinting, which have revealed many new and often unexpected insights into areas such as social systems and migration.

Comprehensive texts on marine mammal biology are few and far between. This is particularly true of books that use a comparative phylogenetic approach, which is based on studying the evolutionary relationships among species and interprets data within an explicitly evolutionary framework. For example, based on the knowledge of phylogenetic relationships among cetaceans and their extinct closest land-bound relatives — the four-legged artiodactyl group, the mesonychids — it is clear that whale locomotion evolved through various stages. In now extinct forms, locomotion advanced from being quadrupedal, to pelvic paddling, and through caudal undulation of feet and tail to the modern condition of vertical movements of the tail only, with the concomitant loss of the hind limbs.

There is great need for a text that takes a strongly evolutionary approach to the study of marine mammals. This book fills that niche. It will be useful for both upper-level undergraduates and graduates and for researchers in marine mammal science. It is well researched, lucidly written and bang up-to-date. I was also impressed by the informed and balanced treatment of current debates, such as that on the phylogenetic relationships within and among cetaceans.

The systematics of marine mammals is relatively well known, largely through molecular phylogenetic investigations over the past ten years. However, questions such as the relationship of the walrus to the other two groups of pinnipeds, the seals and sea lions, or the position of beaked whales to other whales, remain unresolved. Another topic still being debated by molecular phylogeneticists is whether cetaceans should have their own order, or be nested within the artiodactyls, with hippos as their closest living relatives. What is well established is that whales are more closely related to cows and their relatives than they are to donkeys and horses and their perissodactyl relatives. Rates of morphological evolution have apparently been very fast in the whale lineage, rendering their appearance almost totally different from that of their artiodactyl relatives. Yet morphological traits and molecular evidence strongly indicate that donkeys don't belong with cows and whales.

Axel Meyer is in the Department of Evolutionary Biology and Zoology, University of Konstanz, 78457 Konstanz, Germany.

Science in culture

Debate, not didacticism

Shining some scientific light through the anger and the acne at Edinburgh

Sara Abdulla

Amid the angry comedy, fire-eating mime and acne-encrusted Molière of this year's Edinburgh Fringe Festival, there are three plays funded by the Wellcome Trust, "not to be didactic but to engender debate". Three good, thought-provoking, skilfully produced dramas that should send scientists and non-scientists into the night with some meaty ideas to chew on.

Most daring, most theatrical, most creative is Paul Jepson's *The Idiot*, which he and actor Claus Damgaard developed with the help of one of the trust's first Science on Stage and Screen awards in 1998 (worth £35,000 [US\$53,000]). This one-man show is a visceral, troubling, touching exploration of epilepsy based on Dostoevsky's novel.

Impressively, while it stands as an original and exhilarating piece of physical theatre, *The Idiot* is also a theatre-in-education work. It will tour schools (more Wellcome money) augmented by information packs and workshops, helping children better understand, discuss and think about this still-stigmatized condition.

Less thrilling theatrically, but more interesting as science-meets-arts, are two 'straight' plays.

Safe Delivery, another Science on Stage and Screen recipient (to the tune of £36,000) is the likeable result of a collaboration between the venerable playwright Tom McGrath and his geneticist daughter Julie Webb, of the Institute of Child Health in London. Ostensibly about viral gene therapy, it is also about the people beneath the lab coats. "It has struck me as odd since the seventies," says McGrath, "that there are so many scientists in theatre audiences and yet they are never represented on stage."

He addresses that problem in many enjoyable ways. Set in a UK university genetics lab, *Safe Delivery* is a simple drama of human frailty, loss, vanity, ambition, disappointment and love. Some of the basic science is worryingly tabloid — gene therapy straight from test tube to human in three

months without going via animals?

But the characters are otherwise spot on, delicious in their astute, bitchy, fleshy colour. The lecherous, camera-courting media-babe medic with his priorities all askew. The harried prof, so busy writing grants to make sure his empire doesn't lag behind his American collaborators that he doesn't know what's going on in his own cold-room. The eager postdoc, hoarding scant praise. The pretty young PhD student sharing a toast to Barbara McClintock with the embittered, scintillating lecturer she could one day become.

The third of the trio is this year's instalment from Y Touring, beneficiaries of a £300,000 'Creating Debate' grant from the trust for four plays over three years' worth of theatre-in-education touring. *Learning to Love the Grey* is their feisty, if slightly clunky, take on cloning and the conundrums of individual versus collective responsibility that it throws up.

In a contrived and parochial (but ultimately successful) device, writer Jonathan Hall gets in all the necessary information by making this a play about a London woman commissioned to write a play about cloning. One wonders how Y Touring's benefactors took to Hall's astute digs at the plush Public Understanding of Science outfit that his heroine is financially obliged to work with. But no one could disagree with her cry that giving ordinary people information using jargon and condescension is "as bad as not telling us at all".

Structure aside, Hall, who like McGrath has a gratifying flair for dialogue, neatly weaves the myriad issues raised by cloning and stem-cell research into an intelligent and thought-provoking piece. This play will probably have the shortest shelf-life of the three, but UK schoolchildren should benefit enormously from it.

The Trust says it is "very pleased" with these plays. And so it should be.

Sara Abdulla is the senior science writer on the *Nature News Service*.

The Idiot and Learning to Love the Grey are playing at the Pleasance in Edinburgh. Safe Delivery is running at Dynamic Earth. The festival ends on 28 August.



Claus Damgaard in Paul Jepson's play *The Idiot*.

TIM SMITH

Powerful reactions

Nuclear power has taken a meandering route, but it is here to stay.

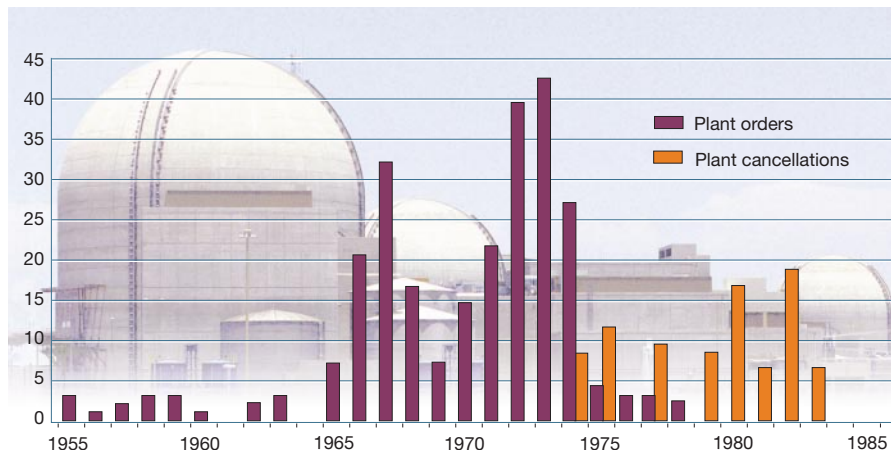
Chauncey Starr

Of all the seminal scientific discoveries, that of nuclear fission in 1939 had uniquely rapid consequences. As in a good mystery story, natural fission had always lurked undetected in the baffling chemistry of radioactive elements. Its discovery opened several doors, the most portentous leading to the first new primary energy source in human history. But a world war made weapons the first intensive US development, causing a massive geopolitical shift. In peacetime, there came radioactive elements for use in medicine and research, and early prototypes of the exciting new energy source, nuclear power.

The task of making nuclear power useful fell to the new US Atomic Energy Commission (AEC), the country's repository of nuclear knowledge. The first half-century of power development was led by physicists and chemists, and by innovative engineers who gradually assumed leadership. Although nuclear power meets many energy needs and is a major option for avoiding greenhouse gases, its growth is constrained by public fear of radiation and by opposition from anti-nuclear environmentalists. Public acceptance is slowly growing as these constraints are discussed and negotiated. In the long term, nuclear power will inevitably be a significant part of our global energy resources, but the social and political trends of today shape its immediate future.

During its first two decades, the AEC benefited from a positive public image and from President Dwight Eisenhower's "Atoms for Peace" promise in 1953. The agency's close ties to the Congress's Joint Committee on Atomic Energy (JCAE) made a team that initiated broad research and

Public acceptance is slowly growing as environmental issues are discussed and negotiated. In the long term, nuclear power will inevitably be a major part of our global energy resources.



Hot and cold: since the mid-1970s, public and government enthusiasm for nuclear power has waned.

century-long plans for nuclear power. Except for the electric utilities, external public and non-scientific policy bodies were generally not invited to participate, for reasons including military security, the arcane sciences involved, and concern with time delays. The AEC/JCAE aimed to help the public understand the fundamentals, and promoted its vision, but treated public input brusquely. This stemmed from the group's belief that its policies were right for society.

It is not surprising, therefore, that the AEC/JCAE was seen in the 1970s as an arm of big government and big industry, and thus a target for the anti-establishment opponents of the Vietnam War. This was the seed of the anti-nuclear dogma that is now uncritically embedded in many public-interest movements. Waning public support for the once-powerful AEC/JCAE moved Congress to dismember it in 1974. In the following five years, the Energy Research and Development Agency absorbed the AEC, and in turn was swallowed up by the Department of Energy, its nuclear budget shrinking at each step.

During this period, the US electric utility industry had a minor role in the first-of-a-kind demonstration plants, sharing the financial risks with the AEC. By the late 1960s, the utilities were confident that nuclear plants were reliable, and placed orders for more. But during the late 1970s and early 1980s many of these orders were cancelled, because national electricity growth slowed, cheap natural gas was released (in 1987) from federal restrictions on power plant use, and government and public support for nuclear power had dwindled. In the 1990s, the energy department became politically anti-nuclear, dominated

by environmentalists, and delayed indefinitely its commitment to store spent nuclear fuel. This remains an unsolved problem for the roughly 100 nuclear power plants now operating in the United States.

Looking back, I believe that there were two strategic mistakes. The first was the AEC's discouragement, verging on disdain, of public participation in planning. In a democracy, elucidation, debate and negotiation with public groups must occur before governments commit to large programmes, so that the public is a responsible party in strategy and risk acceptance. This requires openness, patience and disclosure of objectives — all rare in politicized agencies.

The second misstep was the electric utilities' convenient acceptance of the by-products of weapons materials as a source of fuel. To the utilities, this seemed a cheap option. But this compromise with the military has left the nuclear industry with two problems: a costly hybrid fuel cycle, and a link with weapons in the public mind. The companies in the nuclear enterprise have an annual income of hundreds of billions of dollars; a small fraction of this could have funded reactor concepts and fuel cycles more productively, more cheaply and more quickly.

The relevance of this experience to other new sciences may be limited, but the genetic manipulation of food and drugs is now facing a similar attack from anti-gene environmentalists, who resemble the anti-nuclear movement. It will be interesting to follow that industry's response.

Chauncey Starr is the founder and president emeritus of the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94304, USA.

Worlds of IIF

The facts in the case of doomed, frozen Shankara 3.

Roland Denison

Water is essential for life. Some natural philosophers suggest that life first arose in water, although it is difficult to see how a sufficiently large body of water could ever have existed in our home world's climate to allow the blind clocksmith of evolution to do its work. None exists there now. Perhaps our Sun once shone less harshly, so that lakes could have dotted the planet, forming life's cradle.

Others invoke exotic chemistries to produce life in the polar clouds now trapped and condensed by the few religious fundamentalists who remain after the Diaspora. But it is more likely that primitive life arrived from damper places, on meteorites or other stellar voyagers.

All this makes Shankara 3 the more horrific. That a world should be so profligate with such a precious resource, allowing it to coat 78% of its surface, albeit as ice, seems as near heresy as any rational atheist can devise.

Of course, planets rich in ice are nothing new. Indeed, the Harding drives that powered the early years of the Diaspora were fuelled almost exclusively by water mined from the satellites of gas giants. Equally, Shankara 3 was not the first extrasolar planet found to harbour life. As with our other encounters with alien life, it is mainly unicellular and based on biochemistry similar to our own, although not identical. Shankaran

life is based on nucleic acids, although the actual bases and their pairings are different. This genetic material encodes polypeptides whose amino acids have a little less sulphur and more nitrogen than ours, as well as the curiosity of an imino acid in the minimal alphabet.

But Shankara 3 also supports multicellular organisms. They populate the equatorial regions where temperatures can occasionally creep up close to freezing point. We have named the dominant forms 'blister-bugs' for their curious use of the ice when breeding. Where the male excretes its sperm onto the ice a small 'blister' of liquid water is formed and remains for several days. Into this the females lay their eggs to be fertilized, and here the young grow to adulthood.

The survival of these blisters at sub-freezing temperatures is due to a single polypeptide, Ice-Inhibiting-Factor(IIF)-56, which makes up nearly 40% of the bugs' seminal fluid. IIF-56 binds the encroaching crystals of ice, buying the immature bugs time to grow. But IIF-56 is not indefinite proof against the ice. There always comes a point when the temperature drops sufficiently to allow the ice to prevail, trapping any laggardly bugs in permanent youth.

When water mining began, we found that these were but the last vestiges of a rich fauna that had once swarmed on Shankara 3. Perfectly preserved in the ice is a wealth of exotic beasts, all long dead. All these creatures produced IIFs: the planet's ice is not so much

made of frozen salty water as of IIFs excreted by the animals that lived in it: the vast tracts of ice must once have been open water, a global ocean.

Despite the prevalence of IIFs, phylogenies of Shankaran life based on these polypeptides bear no relation to those based on other characters. Even estimates of the age of life on the planet are different. It is as if the IIFs have spread like an infection so recently that mere hundreds of generations can have passed before the creatures' catastrophic demise.

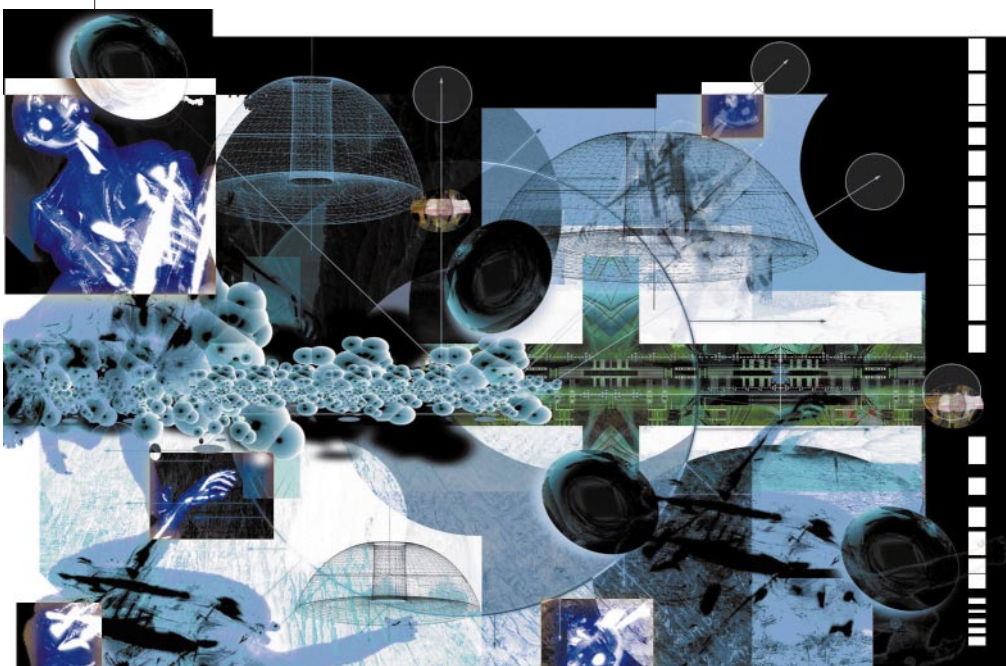
And then the mining rigs found the cities. Vast domes on the rock floor beneath the ice enclosed entire advanced communities: buildings and transport, energy distribution systems and waste disposal routes. Some are shattered but a few intact domes remain, still holding back the tyranny of ice. Inside lie the dome-builders — hundreds of thousands of them in some of the larger domes — frozen in the instant of their deaths going about their normal lives. Fragile bipedal creatures, they are the only life on Shankara 3 that appear to lack IIFs: creatures who could never have survived outside the artificial bubbles they had built.

How they came to live beneath an expanse of water we cannot tell. Perhaps at one time more of Shankara 3 was dry, until a rise in sea level forced these creatures to adopt a subaquarian existence. All we know is how they were living when the ice literally froze them in their ultimate activities. But what if the disaster was of their own making?

When we became aware that our home world would soon fail to support us, we left. Perhaps the Shankarans could not bear to desert the aqueous riches that bathed their planet. When they learned of the imminent coming of the cold did they use their technologies to forestall it? Perhaps the rapid appearance of IIFs in the biosphere was engineered. At first they would have only altered crops to bear IIFs and so better survive colder conditions, but then more and more creatures were modified, subjugating their planet's ecology to build a bulwark against the ice.

Did they realize that the longer they held apart the jaws of this trap, the swifter its eventual closure would be? Finally the ice could no longer be repelled: like the blister-bug's young they were caught for ever in its deadly grip, leaving the planet's uniquely large and solitary satellite to orbit their tombs.

Roland Denison (R.Denison@yahoo.com) works at the XVth Directorate of the European Commission in Brussels. He is best known as the creator of Giles Coppice, the alternative-realities private detective.



Crop strength through diversity

Martin S. Wolfe

In conventional farming, single varieties of crop plants are grown alone. But mixing varieties may be a better option: several rice strains, planted together on a large scale, are more resistant to a major fungal disease.



MICHAEL S. YAMASHITA/CORBIS

HOIT STUDIOS

Attempted solutions to the problems caused by modern agriculture, such as the overuse of fertilizers and pesticides, are usually expensive and often lead to new problems. But this need not be so, as Zhu and colleagues show on page 718 of this issue¹. By growing a simple mixture of rice (*Oryza sativa*) varieties across thousands of farms in China, they restricted the development of rice blast — the most significant disease of rice, caused by a fungus — to levels that are both acceptable and require no treatment with fungicide. This approach is a calculated reversal of the extreme monoculture that is spreading throughout agriculture, pushed by new developments in plant genetics.

Until about 100 years ago, monoculture was practised only at the level of species, with, for example, wheat, maize or rice becoming dominant in different climatic regions. Monoculture has since expanded to different levels, reducing the numbers of species, of varieties within species, and particularly of genetic differences within varieties. Monoculture is convenient: it is easier to plant, harvest, market and identify one variety of crop than several.

But there is a problem. If, for example, all the rice plants in a field are identical, a pathogenic fungus able to attack one plant has a potentially unlimited opportunity to spread throughout the field. At the moment, the solution is either to breed resistant varieties or to develop new fungicides. But the limitless potential for pathogen spread in monocultures leads to rapid selection of pathogens that can overcome resistant crop varieties

Figure 1 The main disease of rice (rice blast, pictured inset) spreads more slowly in mixtures of rice varieties than in monocultures, as Zhu *et al.*¹ discover in their large-scale experiments in China.

and survive in the presence of fungicides. Continual replacement of crops and fungicides is possible, but only at considerable cost to farmer, consumer and environment.

A different approach is to reverse the tide of monoculture by growing several pathogen-resistant varieties as a mixture within a field. Darwin² knew that mixtures of wheat are more productive than single varieties, but explanations for this phenomenon were lacking. It later emerged that mixtures restrict the spread of pathogens and, as a consequence, of disease. The explanation for this phenomenon is complex. The presence of several varieties in a mixture provides a physical barrier to the spread of fungal spores among the plants of one variety. But this is not the only explanation. For example, there is an immunization process among mixed plants. If a form of pathogen that is unable to infect a plant attempts to do so, the plant's disease-resistance mechanisms are activated in the part of the plant affected. Any genetically different spores that would normally be able to infect the plant fail to do so if they try to invade at the same place.

As Zhu *et al.*¹ point out, the net result is a damping of the development of epidemics within the field, with an increase in the complexity of the pathogen population, which may also slow the adaptation of the pathogen to the

mixture³. This is because there may be competition among individual pathogen genotypes that are well adapted to specific varieties in the mixture, and those that thrive on different combinations of varieties but are less specialized. Using different mixtures of varieties in different fields in different years could slow down adaptation of the pathogen even more.

Zhu *et al.*¹ sought to answer one main question: if we can slow down the development of epidemics in one field, what happens if we greatly increase the area of mixed varieties? Will the damping effect multiply across fields? The answer was a clear 'yes'. But first the authors had to persuade all the rice farmers in a large area — within the Yunnan Province, China — that they should grow a particular mixture of rice varieties. The effectiveness of the response from the rice, and from the farmers, thousands of whom participated, was such that it was relatively simple to increase the size of the experiment further in subsequent years. The level of rice blast (Fig. 1; caused by the fungus *Magnaporthe grisea*) was hugely decreased in the target areas, and the farmers stopped using fungicides. This deceptively simple experiment deserves wide attention, partly because of the principle that it illustrates, and partly because it may never be repeated on such a scale.

Some important questions could not be

tackled in this study. The experiment was designed to look at a single major pathogen, the fungus that causes rice blast. But, because the same principles apply to many plant pathogens⁴, it is possible to show that several diseases can be restricted in one crop mixture. For example, during studies for the Elm Farm Research Centre, Hamstead Marshall, Berkshire, UK (see ref. 5), I have recorded the simultaneous restriction of at least three observable diseases in mixtures of wheat varieties relative to single components of the mixtures. There is also evidence that mixtures can buffer against unpredictable abiotic variables, such as cold winter temperatures⁶. Indeed, it is likely that the stability of yields from variety mixtures over different environments³, compared with yields from their components grown as monocultures, results partly from combined restriction of biotic and abiotic stresses.

So why is the mixture approach not used widely? Is it just too simple, not making enough use of high technology? One reason has been concern among farmers and end-users about the quality of the product of the mixtures relative to that of pure varieties: mixtures are said to be unpredictable in terms of quality and ease of harvesting. In practice, such concerns appear to either evaporate or be easily dealt with, as Zhu *et al.* show. In their case, for example, harvesting by hand — a practice common among rice farmers in Yunnan Province — ensured that rice varieties with different qualities could easily be separated and retained for their individual markets. There is also evidence⁷ that mixtures can be designed not only to provide significant disease restriction, but also to improve

product quality by combining complementary characters and providing stability.

Variety mixtures may not provide all the answers to the problems of controlling diseases and producing stable yields in modern agriculture. But their performance so far in experimental situations merits their wider uptake. More research is needed to find the best packages for different purposes and to breed varieties specifically for use in mixtures. And so far researchers have looked only at mixtures of varieties. Mixtures of species provide another layer of crop diversity, with half-forgotten advantages waiting to be exploited in contemporary approaches^{8,9}. It is widely recognized, for example, that high-yielding mixtures of grains and legumes (grass plus clover, maize plus beans, and many other combinations) can restrict the spread of diseases, pests and weeds¹⁰. At the same time, such mixtures can provide near-complete nutrition for animals and humans alike, without recourse to expensive and uncertain forays into genetic engineering. ■

Martin S. Wolfe is at Wakelyns Agroforestry,

Fressingfield, Suffolk IP21 5SD, UK.

e-mail: wolfe@wakelyns.demon.co.uk

1. Zhu, Y. *et al.* *Nature* **406**, 718–722 (2000).
2. Darwin, C. *The Origin of Species by Means of Natural Selection* 6th edn (Murray, London, 1872).
3. Lannou, C. & Mundt, C. C. *Plant Pathol.* **45**, 440–453 (1996).
4. Garrett, K. A. & Mundt, C. C. *Phytopathology* **89**, 984–990 (1999).
5. Finckh, M. R. *et al.* Population Studies of Airborne Pathogens of Cereals (Final Report COST 817, Working Group on Cereal Variety Mixtures, in the press).
6. Maillard, A. & Vez, A. *Rev. Suisse Agric.* **15**, 195–198 (1983).
7. Newton, A. C. & Swanston, J. S. *Annual Report 1998/99* 55–59 (Scottish Crop Res. Inst., 1999).
8. Pimm, S. L. *Nature* **389**, 126–127 (1997).
9. Tilman, D. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
10. Jackson, L. E. (ed.) *Ecology in Agriculture* (Academic, San Diego, 1997).

Artificial noses

Picture the smell

Ingemar Lundström

There is a growing interest in ‘soft’ measurement techniques that measure a particular quality of a sample rather than the quantities of individual properties making up this quality. This type of information gathering mimics the human senses, and has led to the development of ‘electronic noses’ for environmental monitoring, medical testing, and food and drink production. In the most sophisticated systems a unique chemical fingerprint can be generated by an array of sensors and then identified by pattern-recognition techniques as the smell of a rose, for example. On page 710 of this issue¹, Rakow and Suslick suggest that human vision may soon become an important part of what is now known as artificial olfaction.

Attempts to measure odours with electronic instruments² were made as early as the

1950s, but the modern field of artificial olfaction began in 1982 with the work of Persaud and Dodd³, who used a small array of gas-sensitive metal-oxide devices to classify odours. There has since been a steady increase in the number of systems using chemical sensor arrays. The success of artificial olfaction depends not only on the development of new sensor technologies, but also on the availability of powerful pattern-recognition software. This is particularly important for sensor arrays that produce a composite response.

Human vision is probably the most efficient pattern-recognition system available in terms of versatility and speed. Its ability to quickly observe and draw conclusions from changes in the observed images has not been superseded by man-made pattern-recognition systems. It has long been recog-

nized that the most natural way to represent large amounts of data is as an image.

Data from chemical sensor arrays are often presented as images of various types. The most common approach is to cluster data into a two-dimensional image using statistical methods for data reduction and interpretation, such as principal-component analysis. Several ways of creating a visual signature from complex gas mixtures have been suggested — for example, a plot of the different sensor responses in a polar diagram (Fig. 1a).

Ten years ago, our group developed a device in which the properties of a surface coated with catalytic metals change when the surface interacts with different gases^{4,5}. A light pulse scanned across this surface converts the chemical responses into electrical currents, which are then used to generate pixelated images that are quite different for different smells.

A few years ago, another breakthrough in the imaging of smells came with the development of optical-fibre bundles as chemical sensor arrays^{6,7}. In this system, each fibre is coated with a combination of a dye and one of several polymers to give different fluores-

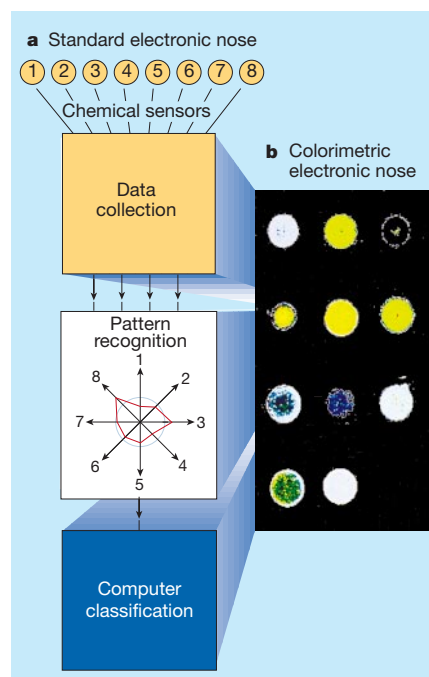


Figure 1 Smelling by colour. a, A typical ‘electronic nose’ consists of an array of chemical sensors with overlapping selectivity profiles for the smells (gas mixtures) to be measured. This is followed by data collection, a pattern-recognition routine (such as the polar diagram shown), and eventually a computer-based decision. b, According to Rakow and Suslick¹ the colorimetric changes of an array of metalloporphyrins upon exposure to organic vapours can replace, with the help of the eye and brain, the various systems used for odour classification. The image here shows an example response from their device to a mixture of 2-methylpyridine and trimethylphosphite vapours.

Microbiology

Lipid lunch for persistent pathogen

William Bishai

cent responses to a gas mixture, thereby producing a pattern of coloured circles at the other end of the fibre bundle. This pattern can be further processed to yield a computer-generated 'olfactory image' of the gas mixture. Such 'olfactory cameras' work in a similar way to biological olfaction in that thousands of odours are detected by a much smaller set of receptor cells (or sensors), which then send signals to the brain, producing a spatially resolved 'neuronal fingerprint'. In this way the natural olfactory system combines high selectivity (for many different odours) with high sensitivity (for trace levels of certain odours).

The 'smell-seeing' device of Rakow and Suslick¹ is based on the colour changes that occur in gas-sensitive metalloporphyrin dyes. Many odour sensors are not able to detect the most toxic vapours, but these compounds bind easily to metalloporphyrins, causing a simple colour change. By fixing an array of different metalloporphyrins in silica gel, the authors are able to produce unique, directly visible, colour patterns when the arrays are exposed to organic vapours, such as ethers (Fig. 1b). They obtain unique 'colour fingerprints' of vapours down to concentrations of a few hundred parts per billion. Although this is competitive with most other chemical sensors, it is still not as sensitive as a natural nose. The commercial viability of this and other devices will also depend on their sensitivity to the background level of humidity, which disrupts many sensors. Rakow and Suslick report that their device is not affected by water vapour.

It is intriguing to be able to identify different smells by eye. Such a system could be used to monitor levels of insecticides in the environment or to sniff out bacteria causing infections. The replacement of pattern-recognition routines and computer-made decisions with the eyes and brain of an experienced (trained) operator will have advantages in many other situations. For example, the screening of antimicrobial peptides to identify replacements for existing antibiotic drugs may soon benefit from a colorimetric sensor array based on lipid vesicles⁸. We have not yet seen the last development in systems to detect specific chemical interactions and, in particular, to 'see the smell'.

Ingemar Lundström is in the Division of Applied Physics, Linköping University, SE-581 83 Linköping, Sweden.
e-mail: ingemar@ifm.liu.se

1. Rakow, N. A. & Suslick, K. S. *Nature* **406**, 710–713 (2000).
2. Gardner, J. W. & Bartlett, P. N. *Electronic Noses: Principles and Applications* (Oxford Univ. Press, 1999).
3. Persaud, K. C. & Dodd, G. H. *Nature* **299**, 352–355 (1982).
4. Lundström, I. *et al.* *Nature* **352**, 47–50 (1991).
5. Winquist, F. & Lundström, I. *Tech. Rundschau* **19**, 54–56 (1993).
6. Dickinson, T. A., White, J., Kauer, J. S. & Walt, D. R. *Nature* **382**, 697–700 (1996).
7. Walt, D. R. *Acc. Chem. Res.* **31**, 267–278 (1998).
8. Kolusheva, S., Boyer, L. & Jelinek, R. *Nature Biotechnol.* **18**, 225–227 (2000).

The bacterium *Mycobacterium tuberculosis* is a bit of an oddity among micro-organisms that invade the human respiratory tract. Most of them, such as *Streptococcus pneumoniae*, enter their host by colonizing the unsterile upper respiratory tract, waiting for a breach in host defences, and then mounting a burst of replication. In contrast, *M. tuberculosis* invades the sterile lungs, avoiding the clearance mechanisms there, and establishes a niche within lung tissue. It then slows or stops multiplying while waiting for the opportunity—afforded by a weakening of the immune system—to produce disease. In this poorly understood state, known as latent *M. tuberculosis* infection¹ (Fig. 1), living *M. tuberculosis* may remain in the body for decades without causing the symptoms of tuberculosis. Defining this unique host–pathogen relationship is a pressing challenge: one in three individuals worldwide have latent *M. tuberculosis* infection, with its 5–10% lifetime risk of progression to active disease.

On page 735 of this issue², McKinney *et al.* show that a biochemical pathway called the glyoxylate shunt is important for the long-term survival of *M. tuberculosis* within mouse tissues. Their results may have implications for the treatment of, and vaccination against, this persistent disease.

Like all bacterial pathogens, *M. tuberculosis* needs to adapt during infection. Years ago, important biochemical changes were identified in mouse-grown *M. tuberculosis*^{3,4}. After developing an *in vitro* model of latency that involves gradual oxygen withdrawal, Wayne^{5,6} noted that 'glyoxylate shunt' enzymes were activated during the metabolic downshift that accompanies oxygen withdrawal⁷. This model was based on two ideas: that lung granulomas—nodular scars containing bacteria and host-cell debris at their centres surrounded by a ring of immune cells—are the sanctuary for latent *M. tuberculosis*, and that oxygen depletion may trigger bacterial adaptation to the latent state.

Isocitrate lyase, the subject of the study by McKinney *et al.*², is one of the glyoxylate-shunt enzymes activated in Wayne's model. Isocitrate lyase and malate synthase together form the glyoxylate shunt, which bypasses the CO₂-generating steps of the tricarboxylic acid (TCA) cycle—the metabolic pathway by which acetate is oxidized to generate ATP. The net result of the glyoxylate shunt is the consumption of two molecules of acetyl CoA to generate one molecule of succinate. Lipids are a source of acetyl CoA; succinate is a precursor for the synthesis of glucose. So the glyoxylate shunt allows *M. tuberculosis*

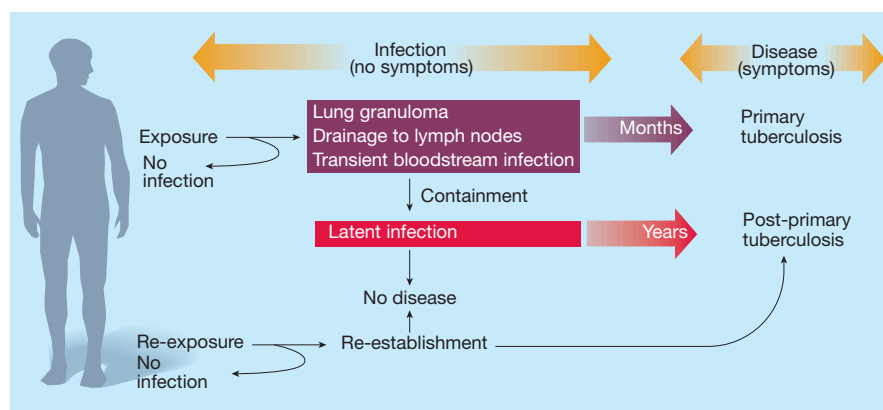


Figure 1 Tuberculosis in humans. This complex disease has several steps, with two forms of progression to active disease—primary and post-primary. *Mycobacterium tuberculosis* infects the lungs, where it grows within macrophages (immune cells that ingest foreign material) during acute infection. An immune response follows, characterized by lung granulomas—nodules containing bacteria and host-cell debris, surrounded by a ring of immune cells. Before full immunity and containment occur, the bacteria probably drain into the regional lymph nodes and the bloodstream. Primary tuberculosis results when acquired immunity fails to contain this initial infection. Successful containment is associated with a failure to eradicate all viable bacteria; this results in a state of latent infection. The location of bacteria during latent infection remains controversial, but old lung granulomas, lymph nodes or remote body sites are possibilities. The results of McKinney *et al.*² indicate that late-stage *M. tuberculosis* may convert lipids into carbohydrates through the glyoxylate-shunt pathway. So the latent bacteria may reside in an environment—perhaps lung granulomas—in which carbohydrates are limited but lipids are available.



100 YEARS AGO

It was mentioned in these columns some time ago that the wire fencing of great sheep farms in some parts of Australia was used as telephone wires. A recent report from H. M. Consul at Philadelphia states that this system of communication is being employed by farmers between the towns of Anderson, Pendleton and Ingalls, in Indiana. The top wire of a barb-wire fence is used as a conductor, the continuity of the line being assured by special devices at highways and railway crossings. The line is fourteen miles in length with five stations... Local farmers state that they have used the "fence-line" to converse with friends eight miles distant, and this at a time when the fence posts were still saturated with the morning dew, a condition under which the line is supposed to work with least satisfaction. It is stated that the line has been such a practical success that the farmers of the neighbourhood are organising companies for the purpose of placing themselves in telephonic communication throughout the whole district.

From *Nature* 16 August 1900.

50 YEARS AGO

Dr. Pablo Osvaldo Wolff, a member of the Expert Committee of the World Health Organisation on Habit-forming Drugs, has written a valuable pamphlet entitled "Marihuana in Latin America, the threat it Constitutes"... Hemp, says Dr. Wolff, has innumerable common names in various countries, and he relates that the hashish of the Middle East, the kif of North Africa..., the dagga of South Africa, the anascha of the U.S.S.R., the esrar of Turkey and Persia, the marihuana of Spanish-speaking America and the United States, and the maconha of Brazil are all names for the same substance, which has other names as well. Dr. Wolff discusses its uses, general effects, the value of experimental observations made with it, its psychopathic manifestations and relationship to crime and delinquency, the medico-legal problems that it creates and the use of it in the Latin American countries. After reading his pamphlet, it is indeed difficult not to agree with him that, while opium "cures and punishes like a god", there is no reason or excuse for the use of marihuana, which "has been closely associated since the most remote time with insanity, with crime, with violence, and with brutality".

From *Nature* 19 August 1950.

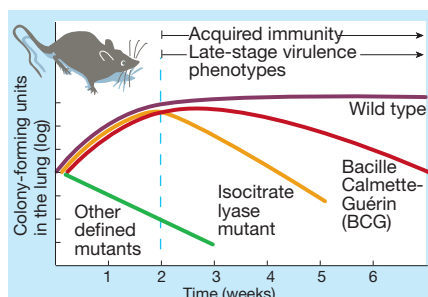


Figure 2 Tuberculosis in the mouse. Mouse tuberculosis involves an initial rise in bacterial number in the lungs (acute infection), followed by a plateau in bacterial burden as a result of acquired immunity. Months later, mice die of progressive inflammation and destruction of lung tissue. The plateau phase (or late stage) may be equivalent to the latent infection phase in human tuberculosis (Fig. 1). McKinney *et al.*² find that *M. tuberculosis* with a mutation in the isocitrate lyase gene grow normally in mouse tissues for the first two weeks of infection, but then show a late-stage defect and are cleared from the lungs. Bacille Calmette–Guérin (BCG), an undefined mutant used as a 'live attenuated' vaccine, shows a similar pattern.

and other bacteria to synthesize carbohydrates from fatty acids. When the TCA cycle and the glyoxylate shunt run simultaneously, fatty acids can supply cells with both energy (through the unabridged TCA cycle) and precursors for carbohydrate synthesis (through the glyoxylate shunt).

McKinney *et al.*² generated an *M. tuberculosis* strain with a mutation in the isocitrate lyase gene that rendered it unable to function. They then tested the ability of the normal and mutant bacteria to infect mice. The progression of tuberculosis in mice (Fig. 2) is slightly different to that in humans. In mice, an initial increase in the number of bacteria in the lungs (the acute phase) is followed, from about two weeks after infection, by a plateau in bacterial abundance. The plateau phase, or late-stage infection, may be roughly equivalent to the latent phase in human tuberculosis. However, in mice the plateau phase is multibacillary (the bacteria are easy to isolate and culture in large numbers), whereas latent infection in humans is paucibacillary (that is, bacteria are difficult if not impossible to find by staining or culturing).

McKinney *et al.* found that the mutated isocitrate lyase gene had no effect on bacterial number in mice during the acute phase. But from two weeks onwards, rather than maintaining a plateau, the number of mutant bacteria in the lungs decreased, indicating that the mutant was being cleared from the lungs (Fig. 2). The need for the isocitrate lyase gene for bacterial survival during late-stage infection indicates that, at two weeks after infection, a change of environment may occur that requires the bacteria to alter their diet from carbohydrate to lipid.

This shift coincides with the onset of acquired immunity against *M. tuberculosis* and the start of granuloma development. In essence, as it becomes important to hunker down to persist within the host, *M. tuberculosis* may begin to dine directly on the lipid-rich host-cell debris in evolving granulomas. In the absence of an external supply of carbohydrates in this new environment, the glyoxylate shunt may supply the sugar precursors required for assembly of the carbohydrate-rich mycobacterial cell envelope.

The three-dimensional structure of *M. tuberculosis* isocitrate lyase has also been reported this year⁸. Several inhibitors of isocitrate lyase were already known, and together these reports indicate that drugs designed on the basis of the structure of isocitrate lyase might lead to new forms of treatment for latent infection. Our current arsenal of anti-tuberculosis drugs, developed over 30 years ago, is plagued by low potency and high toxicity. But, with today's technology, it may become possible to identify potent agents that act against both active and latent *M. tuberculosis*. Indeed, such a drug has been reported recently⁹.

The work of McKinney *et al.* has implications beyond the treatment of tuberculosis, however. Most mutants of *M. tuberculosis* that have been defined so far are cleared during the acute stage of mouse infection. The new work brings into focus another category of *M. tuberculosis* virulence genes — those required for late-stage infection. Other recent studies^{10,11} also support the idea that there is a category of mycobacterial genes that are needed for late-stage survival. This concept may affect anti-tuberculosis vaccination strategies. When tested as live, attenuated vaccines, acute-stage *M. tuberculosis* mutants (which are cleared immediately from animal tissues; Fig. 2) offer no advantage over the currently used, but only partly effective, vaccine bacille Calmette–Guérin (BCG). Long-term protection may require the presentation to the immune system of bacterial antigens expressed in later stages of infection by *M. tuberculosis*¹². So studies of strains attenuated in late-stage infection, such as that produced by McKinney *et al.*, may lead to an improved tuberculosis vaccine.

Despite these advances, fundamental questions about latency linger. Do latent bacteria replicate? How do they escape immune surveillance? Where do they reside? And do models of latency *in vitro* and in mice reflect the situation in humans? The late-stage failure of the isocitrate lyase mutant to persist in mouse lungs fits nicely with the hypothesis^{5–7} that latent *M. tuberculosis* reside in lipid-rich, oxygen-poor granulomas. But the results of human autopsies early in the twentieth century indicated that viable bacteria may reside outside granulomas, in portions of lung tissue unaffected by disease¹³. Late-stage mutants (such as the

isocitrate lyase mutant) offer a new tool with which to investigate these issues. Although much remains to be learned about latent *M. tuberculosis* infection, the identification of late-stage mutants represents a milestone in our efforts to demystify mycobacterial latency. ■

William Bishai is in the Departments of International Health and Medicine, Johns Hopkins University, 615 N. Wolfe Street, Baltimore, Maryland 21205, USA.
e-mail: wbishai@jhsp.edu

1. American Thoracic Society. *Am. J. Respir. Crit. Care Med.* **161**, S221–S247 (2000).

2. McKinney, J. *et al. Nature* **406**, 735–738 (2000).
3. Segal, W. & Bloch H. *J. Bacteriol.* **72**, 132–141 (1956).
4. Segal, W. & Miller W. T. *Proc. Soc. Exp. Biol. Med.* **118**, 613–616 (1965).
5. Wayne, L. G. *Infect. Immun.* **17**, 528–530 (1977).
6. Wayne L. G. & Hayes, L. *Infect. Immun.* **64**, 2062–2069 (1996).
7. Wayne, L. G. & Lin, K. Y. *Infect. Immun.* **37**, 1042–1049 (1982).
8. Sharma, V. *et al. Nature Struct. Biol.* **7**, 663–668 (2000).
9. Stover, C. K. *et al. Nature* **405**, 962–966 (2000).
10. Ramakrishnan, L., Federspiel, N. A. & Falkow, S. *Science* **288**, 1436–1439 (2000).
11. Glickman, M. S., Cox, J. S. & Jacobs, W. R. Jr *Mol. Cell* **5**, 717–727 (2000).
12. Hondalus, M. K. *et al. Infect. Immun.* **68**, 2888–2898 (2000).
13. Opie, E. L. & Aronson, J. D. *Arch. Pathol. Lab. Med.* **4**, 1–21 (1977).

Earth systems

Feedback on Gaia

Jim Gillon

Until a few years ago, self-respecting scientists best avoided making any reference to the Gaia hypothesis — as one delegate fondly recalled at a meeting held in June*, he was once warned that publishing an article with the word ‘Gaia’ in the title could severely damage, if not ruin, a scientific career. Not surprisingly, dissenters still remain, given some of the quasi-religious interpretations of Gaia, such as the notion of the Earth as a living organism. With a name that is Greek for ‘Earth goddess’, maybe this was inevitable. But such conferences show that James Lovelock’s theory¹ of the biotic regulation of Earth has now emerged with some respectability following close scrutiny by the biogeochemical community.

The classic Gaia hypothesis holds that, by introducing feedbacks on climate in particular, life on Earth can regulate and stabilize its environment, possibly indirectly but possibly for its own benefit. One such mechanism proposed by Lovelock^{2,3} is the global thermostat stemming from marine algae that produce dimethyl sulphide, a volatile cloud-seeding chemical. The thinking here is that warmer temperatures lead to greater algal growth and release of dimethyl sulphide into the atmosphere; this stimulates cloud formation, increasing the reflection of radiation back into space and cooling the planet. Some estimates⁴ suggest that in today’s world such cooling might be as much as 4 °C.

In a vastly different early Earth — during the Archaean, the time before 3,000 million years ago — other climate-policing organisms may have been operating. One suggestion at the meeting (J. Kasting, Penn State Univ.) was that, during the Archaean, when the Sun was 30% dimmer and the Earth much cooler, methane-generating bacteria might

have been responsible for warming the planet to a more hospitable temperature as methane, a strong greenhouse gas, accumulated in the atmosphere. At higher concentrations, however, methane molecules can polymerize and form reflective clouds, in effect creating a global photochemical smog that provides a cooling mechanism to stabilize the climate. Similar processes may have occurred, or may still be occurring, in other methane-rich atmospheres — such as that thought to exist early in the history of Mars, or on Saturn’s moon Titan (the Solar System’s best candidate as an extraterrestrial home for life).

Among the main opponents of Gaia theory are evolutionary biologists, some of whom contend that selection of environment-

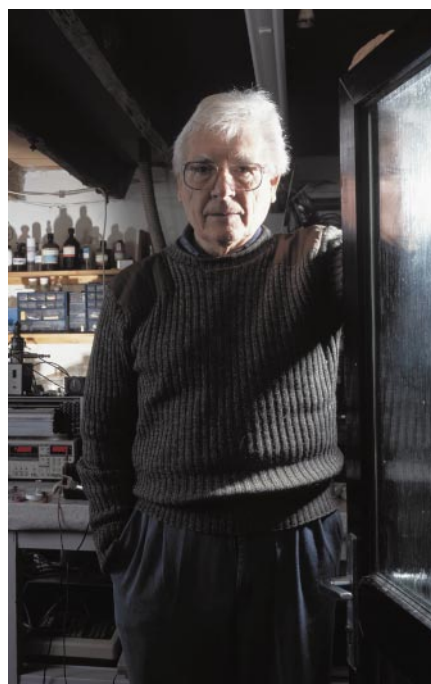
altering traits within organisms is not possible in a darwinian context. Natural selection works at the level of the individual, so how can inherently ‘selfish’ organisms have evolved to change global environment for the common good? One way around this dilemma may come from population biology (D. Williamson, Univ. Liverpool). Self-regulation of populations does not have to be selected for, but is just an emergent property of all individuals trying to maximize their own fitness. Whether this holds for feedbacks on the global scale remains to be seen.

Another criticism of Gaia, voiced frequently at the meeting, centres on the fact that life on Earth has undergone several episodes of near-total extinction — for example, during the periods of almost complete glaciation as a ‘snowball Earth’. This is not the kind of extreme event you would expect from a self-regulating system. Nor indeed are other drastic atmospheric and climatic changes that have followed several of life’s metabolic innovations. One example is the advent of oxygen-generating photosynthesis, which threatened the existence of all anaerobic organisms. Another is the arrival of woody plant tissues, which allowed the huge expansion of vegetation 550 million years ago but induced CO₂-starvation conditions as mineral weathering of rock, stimulated by the action of plant roots, absorbed large quantities of CO₂.

When being swamped by case examples and thought experiments both for and against Gaia, it is impossible to discern a clear-cut consensus. Even so, a point of broad agreement at the meeting was that life can produce feedbacks — usually negative feedbacks — that stabilize the environment and result in long periods of stability. Conversely, the advent of new biochemical processes could, for example, have temporarily introduced positive feedbacks and so have created large phase shifts in the Earth’s conditions. Modelling studies on the interaction between vegetation and climate (E. Eltahir, Mass. Inst. Technol.) show that such two-state systems can exist, with a threshold delimiting recovery to a previous state (stabilizing) or entry into a new phase (destabilizing).

These findings would seem to contradict traditional Gaia theory, with its emphasis on planetary homeostasis. Nonetheless, some delegates argued that global destabilization could be ultimately favourable and so still of a Gaian nature. For instance, although it was detrimental to most anaerobic organisms at the time, the large-scale production of oxygen through photosynthesis could have repackaged the energy that life exploits into a more abundant and ‘user-friendly’ form, allowing the subsequent explosion of aerobic organisms to take place (T. Lenton, Univ. Edinburgh).

In the final analysis, judgement on Gaia



James Lovelock: father of the Gaia hypothesis.

*Second Chapman Conference on the Gaia Hypothesis, University of Valencia, Valencia, Spain, 19–23 June 2000.

theory hangs on the answer to the question: "What have the Gaiaans ever given us?". One response, raised in the summing up (L. Kump, Penn State Univ.), is that an appreciation of Gaia theory has shifted thinking in Earth systems away from cataloguing the fluxes and pools of the Earth's major elements, and towards identifying control systems and feedbacks. Moreover, we have a reminder that there is no harm in taking ideas that were once regarded sceptically and following them through with rigorous

analysis. From these considerations alone, it seems that the general scientific environment has now become hospitable enough for the Gaia hypothesis to last into the future. But how it might evolve is anyone's guess. ■

Jim Gillon is an associate editor at Nature.

1. Lovelock, J. E. *Gaia: A New Look at Life On Earth* (Oxford Univ. Press, 1982).
2. Lovelock, J. E., Maggs, R. J. & Rasmussen, R. A. *Nature* **237**, 452–453 (1972).
3. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
4. Watson, A. J. & Liss, P. S. *Phil. Trans. R. Soc. Lond. B* **353**, 41–51 (1998).

Neurobiology

The shaky trace

Yadin Dudai

Penelope would have been sad indeed had she realized that each time she was reminded of her beloved Odysseus — away from home for so many years after the Trojan war¹ — she could entirely lose her precious memory of him. Fortunately this was also unknown to her pushy suitors. But now Nader and colleagues, writing on page 722 of this issue², have made it public. Recollection, they claim, is a dangerous matter: whenever we bring a memory to mind, it may turn shaky and slip into oblivion.

Most memories, like humans and wines, do not mature instantly. Instead they are gradually stabilized in a process referred to as consolidation^{3,4}. Newly formed memory traces are sensitive to a variety of brain injuries and drugs, but after they have been consolidated they become more resistant to these treatments. Consolidation takes place at many levels of organization and complexity in the brain, and its overall kinetics depends on the type of memory involved. We know most about what happens in individual nerve cells and synapses — the points of communication between neurons — once they have been recruited to consolidate a memory.

The current textbook version, in a nutshell, goes like this. Training modifies proteins at synapses in the neuronal circuit that acquires the new memory. This alters synaptic efficacy and thus the encoding of information in that circuit. But protein molecules survive only for periods of minutes to weeks, whereas many memories are destined to live longer. It seems that at least part of the immunity of memory to this molecular turnover is achieved by training-induced modulation of gene expression in the modified neurons. The new gene products promote long-lasting remodelling of the activated synapses, in a process that involves crosstalk between the synapses and neuronal cell bodies⁵. It takes a few hours for the new pattern of gene expression and the synaptic change to be consolidated. During this time,

the process can be halted by inhibitors of protein synthesis^{5–7}.

This textbook version might tempt one to believe that, for every memorized item, consolidation starts and ends just once. But this view would be naive. Experimental psychologists told us long ago that memory traces are reconstructed with use, and that retrieving a memory involves mingling the representations of the past with the percepts of the present⁸. The study by Nader *et al.*² echoes earlier reports that a consolidated memory can apparently be induced to vanish, provided that the memory is activated shortly before the use of the treatment leading to amnesia^{9,10}. The problem with these early studies was that, because the treatments were applied to the whole brain or even the whole body, and because little was known about the relevant neuronal circuits, the researchers could not target cellular mechanisms in identified memory traces. This has now changed.

Nader *et al.* took advantage of 'auditory fear conditioning' in rats. This works as follows. The rat hears a tone (the conditioned stimulus) in conjunction with a mild foot-shock (the unconditioned stimulus). The electric shock elicits fear (an unconditioned response). After one training session, the tone elicits fear responses, such as freezing, even in the absence of shock (a conditioned response). For readers who are not well versed in the emotional life of rats but do recall the story of Pavlov and his salivating dogs, suffice it to note that the situations are basically similar: both the dogs in Petrograd and the rats in Manhattan had to learn to associate conditioned and unconditioned stimuli. The protocol is therefore aptly dubbed 'pavlovian fear conditioning'. The neuronal circuit underlying pavlovian fear conditioning includes the lateral and basal nuclei of the amygdala. Inhibiting protein synthesis in this brain region immediately after fear conditioning, by infusing the antibiotic anisomycin into this region, blocks long-term fear memory

(that existing more than 24 hours after the training), but not short-term memory².

Knowing all this, Nader *et al.* trained rats in pavlovian fear conditioning, and tested them 24 hours later with the conditioned stimulus but without the unconditioned stimulus (test 1). The rats froze at the sound of the tone. At this point, when the long-term memory trace was expected to be already insensitive to anisomycin, Nader *et al.* injected the antibiotic into the amygdala. A day later, the authors tested the rats again with just the conditioned stimulus (test 2). Surprisingly, these rats showed a marked decrease in the time spent freezing in response to the tone. The same results were obtained even if test 1 took place 14 days after training, making it even more unlikely that the inhibition of protein synthesis in test 1 impaired a late phase of consolidation initiated by the original training. Omitting the conditioned stimulus before administering anisomycin in test 1 left memory intact. So the memory probably had to be retrieved for anisomycin to have its effect. The anisomycin was effective only if administered within a few hours after memory reactivation.

So it seems that fear-associated memories become temporarily labile on retrieval. Why should the brain invest so much energy in the original consolidation and then risk losing the trace by interference each time it is used? One can come up with teleological explanations — for example, that the brain prefers plasticity at the expense of stability — or mechanistic ones, suggesting in-built constraints on the synaptic machinery. But there is still much to do before we can jump to any sweeping conclusions about the cellular biology of memory retrieval. Some unanswered questions relate specifically to this experiment. Did tests 1 and 2 indeed tap the same memory trace? Did anisomycin abolish the original trace, or merely leave it dormant, waiting to be exposed by some smart behavioural protocol? Which cellular mechanisms are perturbed by anisomycin after retrieval, and are they identical to those that produce the original consolidation?

More generally, might these results apply to different types of memory? Previous studies hinted that pavlovian fear conditioning may not be unique in being shaky on retrieval^{9,10}. But even if just a few types of memory must reconsolidate after use, the implications of the results of Nader *et al.*² are remarkable. Consider, for example, the prospect of intentionally recalling the memory of a traumatic experience and then selectively erasing it. What such a possibility would mean for psychoanalysts on the one hand, and poets on the other, is quite a different matter. ■

Yadin Dudai is in the Department of Neurobiology, The Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: yadin.dudai@weizmann.ac.il

1. Homer *The Odyssey* Book I (transl. E. V. Rieu) (Harper, New York, 1991).
2. Nader, K., Schafe, G. E. & Le Doux, J. E. *Nature* **406**, 722–726 (2000).
3. Dudai, Y. *Neuron* **17**, 367–370 (1996).
4. McGaugh, J. L. *Science* **287**, 248–251 (2000).
5. Dudai, Y. & Morris, R. G. M. in *Brain, Perception, Memory. Advances in Cognitive Sciences* (ed. Bolhuis, J. J.) 147–162 (Oxford Univ. Press, in the press).
6. Goelet, P., Castellucci, V. F., Schacher, S. & Kandel, E. R. *Nature* **322**, 419–422 (1986).
7. Davis, H. P. & Squire, L. R. *Psychol. Bull.* **96**, 518–559 (1984).
8. Tulving, E. *Elements of Episodic Memory* (Oxford Univ. Press, 1983).
9. Misanin, J. R., Miller, R. R. & Lewis, D. J. *Science* **160**, 554–555 (1968).
10. Sara, S. J. *Learn. Memory* **7**, 73–84 (2000).

Organic materials

From insulator to superconductor

Philip Phillips

The discovery in the late 1970s that certain types of polymeric material¹ can conduct electricity almost as efficiently as copper wire paved the way for light-emitting diodes, electrical sensors and even rechargeable batteries made entirely from plastics. This success sparked general interest in electronic devices made from cheap organic materials, such as organic superconductors². In a superconductor, electrons bind together to form Cooper pairs that carry current without resistance at sufficiently low temperatures. On page 702 of this issue³, Batlogg and his colleagues describe how to make superconducting transistors from anthracene, tetracene and pentacene. Under normal conditions, these simple organic materials are insulators, but they can be converted into superconductors simply by injecting them with charge and lowering the temperature.

The organic crystals studied by Batlogg and colleagues are composed of three (anthracene) to five (pentacene) linked rings of carbon atoms. As part of a transistor they can conduct electricity, and at low temperatures of 2–4 K their electrical resistance drops to zero. This superconducting behaviour is destroyed by an applied magnetic field. Zero resistance and expulsion of magnetic fields are intrinsic features of a true superconducting state. But given the high superconducting transition temperatures ($T_c = 135$ K) that have been observed in copper-oxide compounds, why should we care about a relatively low T_c in an organic transistor? The answer is simple — these materials are far from being ordinary superconductors and offer a fertile playground for studying the fundamental properties of correlated electrons.

Batlogg and colleagues build their organic superconductors from highly pure single crystals of polyacenes coated with an insulator, typically aluminium oxide. By fixing a gate electrode to the oxide layer and placing source and drain electrodes on the organic crystal, they create a field-effect transistor (see Fig. 1 on page 702). Applying a voltage to the gate electrode triggers a flow of charge between the other two electrodes. Such a device is electrically identical to those made from semiconductors, such as silicon (Si) and

gallium arsenide (GaAs). What is different about the organic devices is that they become superconducting at high charge densities. Superconducting materials have been created from insulators before by doping them with oxygen, as in the case of the copper oxides; but this new type of 'charge doping' greatly broadens the range of materials that can be converted into superconductors.

Superconductivity is generally associated with a bulk material. But in these experiments, electrons cannot penetrate the insulator. As a result, they are confined to move within the thin two-dimensional interface between the organic crystal and the oxide. This is significant because, in the past 20 years, the physics of two-dimensional electron gases has dominated solid-state research, resulting in Nobel prizes in 1985 and 1998. Exotic phases, such as those found in copper-oxide superconductors or in a highly correlated electron liquid state with fractionally charged excitations⁴, are the main reasons why the physics of correlated two-dimensional electrons is such an important problem. Many of these exotic states can now be studied in an organic field-effect transistor — indeed, Batlogg's group has already observed the quantum Hall effect in pentacene⁵.

As with any discovery of superconductivity, the physical mechanism behind the electron pairing is the key issue. It is worth noting that no chemically doped polyacene has ever been observed to superconduct. In this respect, the observation of superconductivity in a fullerene (C_{60} molecule) transistor⁶ is quite different from that in the polyacenes, because C_{60} doped with alkali metals superconducts. Furthermore, it is thought⁷ that doped C_{60} follows the standard picture of superconductivity in which tiny lattice vibrations (or phonons) provide the glue that binds electrons in pairs. The strength of this glue can be inferred from the electron–phonon coupling constant⁸. For a series of organic compounds, including the polyacenes and C_{60} , the electron–phonon coupling constant falls off inversely with the number of π -electrons associated with the carbon rings (Fig. 1).

On the basis of this trend, one expects anthracene to have the highest superconducting transition temperature, T_c , within the polyacene series studied by Batlogg and co-workers. Indeed, this is what they find. But one also expects doped C_{60} to have the lowest, whereas it has a much higher T_c of 33 K. The explanation may lie in differences in their density of electronic states. It is the product of the density of states and the electron–phonon coupling constant that determines, to a large extent, the value of T_c . But an independent measure of the density of states is difficult to achieve experimentally unless thermodynamic measurements are made. Such measurements would be particularly difficult in the polyacene crystals because the electron gas is sandwiched between two insulators.

An experiment that could help determine the mechanism of pairing in the polyacenes (and that is now possible) would be to probe the properties of a pentacene transistor at low temperatures and low electron densities.

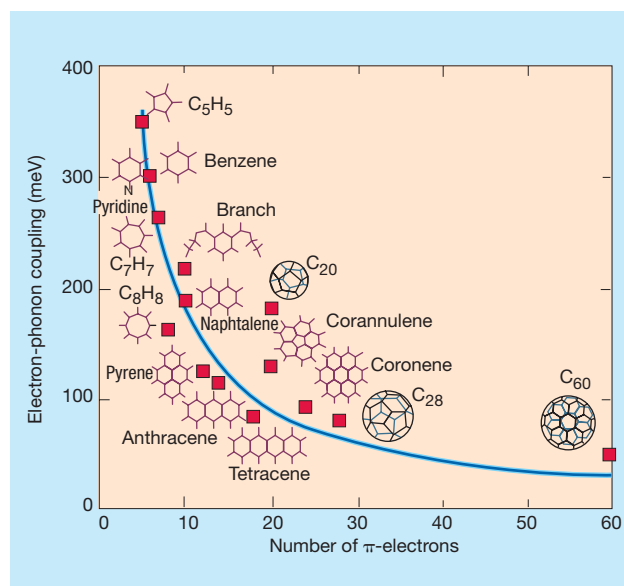


Figure 1 What lies behind the superconductivity of organic materials, such as polyacenes (anthracene) and fullerenes (C_{60})? A clue might be provided by the electron–phonon coupling constant for a series of organic semiconductors is shown as a function of the number of π -electrons associated with their carbon rings. (Redrawn from ref. 8.)

Varying the electron density changes the strength of the Coulomb repulsion relative to the kinetic energy of the electrons. The Coulomb repulsion falls off inversely with distance, whereas the kinetic energy decays as the inverse square of the distance, so the Coulomb repulsion dominates in the low-density regime. In this regime, phonons are too weak to overcome the repulsive interaction between electrons. So the observation of superconductivity in the dilute regime would imply that pairing occurs entirely from repulsive interactions. That is, pairing could be a general organizing principle⁹ (or 'quantum protectorate' to use the language of Laughlin and Pines¹⁰) of a dilute two-dimensional electron gas. Other more complex mechanisms involving disorder¹¹ and time-reversal broken states¹² have also been discussed.

The observation of superconductivity in a pentacene transistor at low electron densities would fuel the current debate over an unexpected conducting state seen in a dilute two-dimensional electron gas in gated semiconductor transistors¹³. There is no consensus on the origin of this state, but the same density regime could be examined in a pentacene transistor simply by changing the gate voltage. If it turns out that superconductivity persists in the pentacene systems at low densities, the question will be why we have not seen superconductivity in two-dimensional electron gases in inorganic semiconductor devices, such as Si and GaAs. After all, if superconductivity is an intrinsic property of a dilute two-dimensional electron gas, then it should not matter what material the transistor is made of. This question may be difficult to answer given the growing evidence¹⁴ that Cooper pairs can also exist in a metallic rather than a truly superconducting state at low temperatures. Nonetheless, the achievements of Batlogg and colleagues offer a new avenue for tuning the magnitude of the electron interactions to study almost any aspect of two-dimensional solid-state physics. ■

Philip Phillips is at the Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1100 W. Green Street, Urbana, Illinois 61801-3080, USA.

e-mail: philip@vfemmes.physics.uiuc.edu

- Chiang, C. K. *et al.* *Phys. Rev. Lett.* **39**, 1098–1101 (1977).
- Little, W. A. *J. Phys. (Paris) Coll.* **44**, 819–825 (1983).
- Schön, J. H., Kloc, C. & Batlogg, B. *Nature* **406**, 702–704 (2000).
- Laughlin, R. B. *Phys. Rev. Lett.* **50**, 873–876 (1983).
- Schön, J. H., Kloc, C. & Batlogg, B. *Science* **288**, 2338–2340 (2000).
- Schön, J. H., Kloc, C., Haddon, R. C. & Batlogg, B. *Science* **288**, 656–658 (2000).
- Gunnarsson, O. *Rev. Mod. Phys.* **69**, 575 (1997).
- Devos, A. & Lannoo, M. *Phys. Rev. B* **58**, 8236–8239 (1998).
- Phillips, P. *et al.* *Nature* **395**, 253–257 (1998).
- Laughlin, R. B. & Pines, D. *Proc. Natl Acad. Sci. USA* **97**, 28–31 (2000).
- Belitz, D. & Kirkpatrick, T. *Phys. Rev. B* **58**, 8214–8217 (1998).
- Zhang, F.-C. & Rice, T. M. <http://xxx.lanl.gov/abs/cond-mat/9708050>
- Sarachik, M. P. & Kravchenko, S. V. *Proc. Natl Acad. Sci. USA* **96**, 5900–5902 (1999).
- Mason, N. & Kapitulnik, A. *Phys. Rev. Lett.* **82**, 5341–5344 (1999).

Cancer

Taking up iodide in breast tissue

Piri L. Welch and David A. Mankoff

After the disaster at the nuclear power plant at Chernobyl, more than 800 children who drank milk from exposed dairy herds developed thyroid cancers¹. The cancers were caused largely by exposure to radioactive fallout (most of which consisted of radioiodide, derived from iodine-131) in grass; the radioiodide was concentrated in the cows' mammary cells, secreted into their milk, ingested by children, and concentrated in the children's thyroid glands, causing cancer. This was a tragic consequence of the essential process of iodide transport into both thyroid and mammary cells. Transport of iodide into thyroid cells is mediated by a transmembrane carrier protein, the sodium-iodide symporter (NIS). Writing in *Nature Medicine*², Tazebay, Wapnir and colleagues show that the NIS is also responsible for iodide transport into epithelial cells of the normal lactating breast and breast cancer cells (Fig. 1). Their results may reopen the 60-year-old discussion of the possible use of iodide transport for detecting or treating breast cancer, and suggest a possible new way of treating thyroid cancer.

Iodinated compounds are essential for the normal growth and development of vertebrates. The thyroid gland actively accumulates iodide from the blood against an electrochemical gradient, and incorporates it into thyroxine and tri-iodothyronine hormones, which control the basal metabolic rates of many cells. It has been known for at least 60 years that the breast also concentrates iodide, secreting it into milk³. Nursing infants synthesize essential thyroid hormones using iodide from breast milk. Active transport of iodide into the thyroid is mediated by the NIS^{4,5}, which simultaneously transports two different ions — sodium and iodide — in the same direction across the plasma membrane of thyroid cells. Tazebay *et al.*² now show that the NIS also concentrates iodide in normal lactating breast ducts.

During pregnancy and lactation, dramatic

changes occur in the normal mammalian breast. During pregnancy, breast enlargement results from hormonally stimulated, rapid proliferation of epithelial cells of breast alveoli and ducts. The synthesis in the pituitary gland of the milk precursor protein prolactin also increases, peaking at birth with production of milk. Suckling further stimulates prolactin synthesis, leading to more milk production. Tazebay *et al.* find that the NIS protein is present in mammary tissue in mice just before birth and during lactation, but that NIS expression is significantly decreased at weaning. If suckling begins again, NIS levels rise, suggesting that NIS expression is hormonally regulated. Experimentally, a combination of oestrogen, oxytocin and prolactin — the hormones present during natural lactation — most effectively induces NIS expression in mammary tissues. In the breast-cancer cell line MCF-7, retinoic acid has also been shown to stimulate NIS expression and radioiodide uptake⁶.

The enormous burst of proliferation of breast epithelial cells during normal pregnancy and lactation is echoed by the abnormal cellular proliferation occurring in breast cancer. As shown by Tazebay *et al.*, NIS is also expressed in the mammary tumours of mice whose cancers are caused by experimental overexpression of the oncogenes *HER2/neu* or *ras*. The authors find that mouse mammary tumours that express NIS — unlike those that do not express NIS, or normal, non-lactating mouse mammary glands — accumulate a radioactively labelled iodine substitute. It was reported some 25 years ago⁷ that human breast tumours (whether benign or invasive) take up more radioiodide than do normal breast epithelial cells. It seems that NIS may be responsible: 87% of the primary invasive human breast tumours analysed by Tazebay *et al.* expressed NIS, but none of the non-lactating breast tissue did.

The specific uptake of radioiodide through the NIS is invaluable in the diagnosis

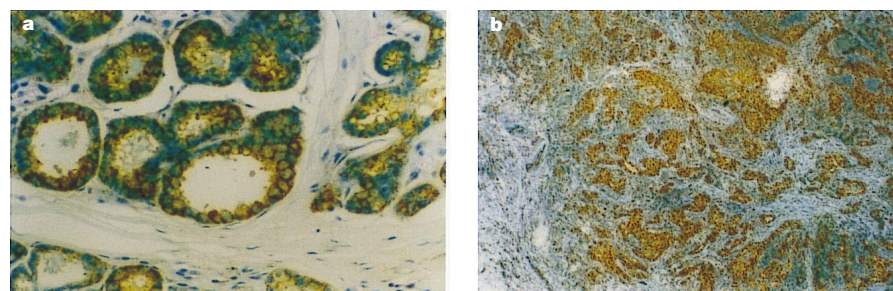


Figure 1 Expression of the sodium-iodide symporter (NIS) in breast tissue. Tazebay *et al.*² have shown that lactating (a) and cancerous (b) breast tissues express the NIS.

and treatment of both benign and malignant thyroid disorders. For example, radioiodide is commonly used to treat overactive thyroid glands, such as those seen in Graves' disease. Thyroid cancer cells generally retain their ability to transport iodide, although to a lesser extent than normal thyroid cells. So, for patients whose thyroids have been removed during treatment for thyroid cancer, whole-body radioiodide scanning is often used to detect cancerous cells of thyroid origin that have migrated (metastasized) to other sites in the body. Often during these clinical procedures, iodide uptake also occurs in other tissues such as salivary gland and lactating breast. This is consistent with the observation² that the NIS is expressed in some non-cancerous tissues near breast tumours. Administration of higher doses of radioiodide is an effective treatment for metastasized thyroid cancer.

The implications of the new findings² for diagnosing and treating breast cancer are uncertain, however. Unlike thyroid-cancer patients, most breast-cancer patients have functioning thyroids. Normal thyroid cells accumulate and retain iodide far more efficiently than do thyroid cancer cells, breast epithelial cells, or any other cell type⁸. The presence of a patient's normal thyroid will pose a challenge to the use of radioiodide to detect or treat breast tumours, as it will sequester nearly all the radioiodide until the thyroid itself is destroyed. Also, it is not yet known what proportion of human breast cancers express functional NIS and hence accumulate iodide. The results of Tazebay *et al.* indicate that the NIS may be incorrectly localized within the cells of some breast tumours and thus incapable of taking up iodide. So, more work is needed before we can know whether uptake of iodide in breast tumours will be of clinical use.

The results have greater implications for thyroid cancer patients. Treating such patients with radioiodide is effective when the radioiodide is retained in the cells; however, metastases can lose their ability to take up and accumulate radioiodide. In this case, thyroid cancers often develop further, and other treatments — such as external beam radiotherapy — are usually ineffective at stopping the disease. So there are intense efforts to try to stimulate such thyroid tumours to 'redifferentiate' and take up radioiodide again. Tazebay *et al.*² have shown that it is possible to use hormones to stimulate the expression of NIS in non-lactating mouse mammary cells. Perhaps it might be possible to stimulate thyroid metastases to express NIS in the same way, and hence to regain the ability to accumulate enough radioiodide for effective treatment. ■

Piri L. Welch is in the Department of Medicine, Division of Medical Genetics, and David A. Mankoff is in the Department of Radiology, University of Washington, Seattle,

Washington 98195, USA.

e-mails: piri@u.washington.edu
dam@u.washington.edu

1. Pacini, F. *et al.* *Acta Paediatr.* **88** (suppl.), 23–27 (1999).
2. Tazebay, U. H. *et al.* *Nature Med.* **6**, 871–878 (2000).
3. Elmer, A. *Iodine Metabolism and Function* (Oxford Univ. Press, London, 1938).

4. Carrasco, N. *Biochim. Biophys. Acta* **1154**, 65–82 (1993).
5. Eskandari, S. *et al.* *J. Biol. Chem.* **272**, 27230–27238 (1997).
6. Kogai, T., Schultz, J. J., Johnson, L. S., Huang, M. & Brent, G. A. *Proc. Natl Acad. Sci. USA* **97**, 8519–8524 (2000).
7. Eskin, B. A., Parker, J. A., Bassett, J. G. & George, D. L. *Obstet. Gynecol.* **44**, 398–402 (1974).
8. Filetti, S. *et al.* *Eur. J. Endocrinol.* **141**, 443–457 (1999).

Global change

Ice sheets by volume

Peter U. Clark and Alan C. Mix

In 1840, Louis Agassiz proposed that vast ice sheets had once covered large areas of northern Europe and North America. Since then, geologists, geophysicists and geochemists have argued about the spatial extent and volume of the ice during the Last Glacial Maximum (LGM) around 21,000 years ago.

On page 713 of this issue, Yokoyama *et al.*¹ present evidence that puts a firm figure on ice volume at the LGM. Growth of continental ice lowers global sea level (referred to as glacio-eustatic sea level), so records of sea-level change can be used to reconstruct changes in ice volume. Using new records from the continental shelf off northern Australia, Yokoyama *et al.*¹ estimate that, during the LGM, the sea level was 130–135 m lower than it is now. From this value, they calculate that the LGM ice volume was $(52 \pm 2) \times 10^6$ km³ more than the present volume of 32×10^6 km³. In combination with other data, this information can be used to address the interplay between ice sheets, the climate system and the solid Earth.

Ice sheets have a huge impact on Earth's climate by rearranging continental drainage paths and by changing the Earth's topography and albedo (the amount of the Sun's radiation reflected from Earth). At the LGM, ice sheets were nearly 4 km high, covered some 13 times more land than they do today (excluding Antarctica, where the ice area did not change appreciably), and lowered the sea level to expose an extra 2.8×10^6 km² of land now under water (nearly equal to the area covered by the additional LGM ice).

Experiments with numerical climate models often target the LGM climate to evaluate climate sensitivity to radiative 'forcing' from such factors as ice albedo, the amount of CO₂ in the atmosphere and variations in the Earth's orbital parameters. The models show that ice sheets contributed as much as 50% of the total forcing². But uncertainties remain. Although the outlines of most of the major ice sheets are now reasonably well known, in some regions the geological record of ice-sheet extent is difficult to interpret — particularly in northern Eurasia^{3,4}. Because

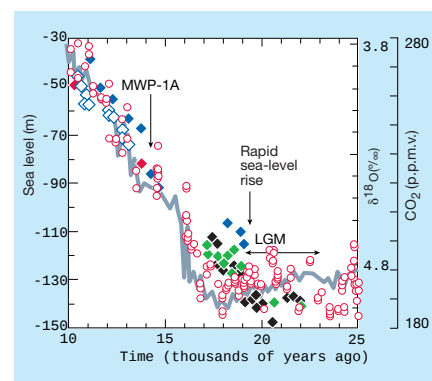


Figure 1 Sea-level estimates for 25,000–10,000 years ago compared to marine oxygen isotope and atmospheric CO₂ records. The rise over this period is taken to be due to glacio-eustatic effects (that is, the retreat of ice sheets). The new data of Yokoyama *et al.*¹ are shown as black diamonds; green diamonds represent data from Barbados corals corrected by Yokoyama *et al.* Other estimates¹² are based on previous eustatic corrections to coral-reef data in Barbados (blue diamonds), Tahiti (red diamonds) and New Guinea (open blue diamonds). Generally, these estimates match the values implied by the oxygen-isotope ($\delta^{18}\text{O}$) record (grey line)¹⁶. But a rapid rise in sea level about 19,000 years ago, identified by Yokoyama *et al.*, precedes the rapid changes in both $\delta^{18}\text{O}$ and atmospheric CO₂ (open red circles)¹⁷. MWP-1A (meltwater pulse 1A) is an interval of rapid sea-level rise first dated in the Barbados record¹⁸. LGM, Last Glacial Maximum¹⁷; p.p.m.v., parts per million by volume.

the ice sheets left little direct evidence of their height, estimates of LGM ice volume have come largely from indirect evidence or from glaciological modelling. The resulting figures, put in terms of equivalent sea-level lowering, have ranged from 80 m to as much as 165 m.

Among the indirect indicators are oceanic oxygen isotopes, expressed as $\delta^{18}\text{O}$ in parts per thousand, preserved in fossil shells. Ice-sheet growth caused a measurable change in the $\delta^{18}\text{O}$ composition of the global ocean⁵, but the relative contributions of other factors (such as temperature and local salinity) to this signal remain poorly understood, and

floating ice shelves would change $\delta^{18}\text{O}$ but not sea level. In another approach, records of local isostatic responses (that is, the rise and fall of the Earth's crust) to changes in ice thickness or water depth can be inverted to estimate ice thickness⁶. However, uncertainties in model parameters, and the brevity of critical time series, mean that several solutions are possible⁷. Finally, modelling of ice volume based on the area occupied by former ice sheets initially assumed that the conditions at the base of ice sheets, which control ice profile, were the same in the past as they largely are now. Yet they may have been very different⁸.

The most direct evidence of LGM ice volume comes from records of lower sea level. But there are two difficulties: finding a well-preserved and dateable record of LGM sea level, and then distinguishing the isostatic from the glacio-eustatic component of the signal. Yokoyama *et al.*¹ addressed these difficulties by dating geological records on the tectonically stable northern Australian continental shelf, and deriving the glacio-eustatic component by accounting for the isostatic effect on the shelf caused by the sea-level rise that accompanied deglaciation.

Their results resolve a long-standing controversy. Furthermore, they suggest that the LGM ice volume was relatively stable for at least 3,000 years, implying that the ice sheets approached isostatic and, perhaps, dynamical equilibrium. Their analysis also indicates that the LGM was terminated by a rapid rise in sea level 19,000 years ago (Fig. 1). Such an abrupt event may record a climatic or other instability that triggered the demise of the ice sheets^{1,8}, and may help to explain the conflicts with ice-volume reconstructions based on inversion of past sea levels⁷.

Useful though they are, the new data do not show how the ice was distributed among individual ice sheets. Once the spatial extent of the former ice sheets in northern Eurasia is resolved^{3,4}, models should converge on some reasonable distribution of ice that sums to 130–135 m of lowered sea level. This is important, as climate models are sensitive to regional details of ice extent and height.

But some significant puzzles remain in reconciling the various data sets. For example, the new sea-level estimates support a change of some 1.4‰ in the $\delta^{18}\text{O}$ budget of the whole ocean. This is consistent with some estimates of deep-sea cooling⁹, but it conflicts with those based on pore-water data¹⁰. Similarly, the isotope-budget data suggest that temperatures in the oceanic warm pools, such as the western tropical oceans, were relatively stable — this runs counter to temperature indices based on strontium:calcium ratios in fossil corals¹¹.

One next step will be to confirm the Australian record using other sea-level archives, such as submerged coral reefs. Corals that grow on the ocean margins can be accurately

dated, and coral records from Barbados, New Guinea and Tahiti provide valuable information on past ice volume. But existing coral records¹² miss both the LGM low sea level and the early jump in level of around 19,000 years ago seen in the Australian record. Combining coral records with the new data, however, suggests that deglaciation was punctuated by two, or perhaps three, brief intervals when rates of sea-level rise exceeded 5 m per century (Fig. 1).

The inferred sea-level jump of 19,000 years ago precedes the first prominent decrease in marine $\delta^{18}\text{O}$ by about 3,000 years (Fig. 1), and if this is correct it indicates a decoupling of the timing or amplitude (or both) of the sea-level and isotope records during this early phase of deglaciation. Such a misfit is too large to be accounted for by ocean mixing¹³. Together with a similar lag in increasing atmospheric CO_2 , these relations seem to require ocean cooling associated with the jump in sea level, or dynamic changes involving the collapse of ice on land compensated by growth of floating ice shelves. Such a shift might displace water in the ocean without modifying the oxygen isotope budget, while helping to keep CO_2 in the deep sea.

What really happened early in deglaciation remains a matter of debate¹⁴. So too does the timing of events as reflected in ice cores and marine sediments. But all of these factors are crucial to understanding the cause of deglaciation, as well as the potential for ice-sheet instability in the future. These issues, among others, will be addressed at a meeting in October this year held under the auspices of the "Environmental Processes of the Ice Age: Land, Oceans, Glaciers" (EPILOG) programme¹⁵. ■

Peter U. Clark is in the Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, USA.

e-mail: clarkp@ucs.orst.edu

Alan C. Mix is in the College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA.

e-mail: mix@oce.orst.edu

1. Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L. K. *Nature* **406**, 713–716 (2000).
2. Hewitt, C. D. & Mitchell, J. F. B. *Clim. Dyn.* **13**, 821–834 (1997).
3. Svendsen, J. I. *et al.* *Boreas* **28**, 234–242 (1999).
4. Grosswald, M. G. & Hughes, T. J. *Polar Geogr.* **23**, 23–54 (1999).
5. Shackleton, N. J. *Nature* **215**, 15–17 (1967).
6. Peltier, W. R. *Rev. Geophys.* **36**, 603–689 (1998).
7. Lambeck, K. *et al.* *Earth Planet. Sci. Lett.* (in the press).
8. Clark, P. U., Alley, R. B. & Pollard, D. *Science* **286**, 1104–1111 (1999).
9. Mix, A. C. & Pisias, N. G. *Nature* **331**, 249–251 (1988).
10. Schrag, D. P. *et al.* *D. W. Science* **272**, 1930–1932 (1996).
11. Guilderson, T. P. *et al.* *Science* **263**, 663–665 (1994).
12. Fleming, K. *et al.* *Earth Planet. Sci. Lett.* **163**, 327–342 (1998).
13. Duplessy, J. C. *et al.* *Earth Planet. Sci. Lett.* **103**, 27–40 (1991).
14. Clark, P. U. *et al.* *Paleoceanography* **11**, 563–577 (1996).
15. Mix, A. C., Bard, E. & Schneider, R. *Quat. Sci. Rev.* (in the press).
16. Mix, A. C. *et al.* in *Mechanisms of Global Climate Change at Millennial Time Scales Geophys. Monogr.* **112** (eds Clark, P. U., Webb, R. S. & Keigwin, L. D.) 127–148 (Am. Geophys. Union, Washington DC, 1999).
17. Sowers, T. & Bender, M. *Science* **269**, 210–214 (1995).
18. Bard, E. *et al.* *Nature* **345**, 405–410 (1990).

Daedalus

Shuffling around

Storage is costly for manufacturers and distributors alike. Indeed, 'just in time delivery' transformed the economics of many companies simply by reducing their burden of storage. Yet in any warehouse, over half the space is always empty — the bays and aisles needed for access to the goods.

In this connection Daedalus recalls the '15 puzzle', the mathematical amusement in which a tray of four units square contains 15 numbered tiles and one space. By repeatedly moving an adjacent tile into the space, you can shuffle the tiles into almost any chosen pattern. (In fact, just half the imaginable arrangements are accessible.) So DREADCO engineers are now inventing a '15-puzzle warehouse'. Instead of a fixed floor, it has a number of mobile platforms running on a grid of rails set out in a square pattern. Each platform carries a tower of pallets bearing the goods to be stored. They completely fill the warehouse, except for one space. Yet that single space allows the platforms to be shuffled indefinitely. Any desired platform can be rapidly brought to the loading bay for the delivery or replenishment of its contents.

The mathematics of the 15-puzzle warehouse (or rather its generalization, the 'NM-A warehouse', where *N* and *M* are the unit numbers of its sides, and *A* is the number of empty spaces needed for fast enough access to the least accessible platform) will be subtle indeed. While DREADCO's hardware team consider the wheels, rails and motors needed to drive the platforms around under computer control, their software colleagues are writing the algorithms to access any platform or succession of platforms in the fewest moves. The final fully automated installation will need no fork-lift trucks to ferry goods around. A simple computer command will rapidly bring any desired sequence of goods or empty platforms to the one portal that handles all transactions. Storage costs will plummet.

The NM-A principle should also usefully apply in other fields. The NM-A car-park, its vehicles packed in a solid rectangular array, will save vast amounts of urban space. A naive owner may be dismayed to see his vehicle zig-zag away into the crowded mass. But once in the matrix it will be impossible to steal or steal from, and only the owner's ticket in the car-park computer will shuffle it back to the exit. But an NM-A library, or even an NM-A supermarket, would probably be intolerably inconvenient. Customers for these amenities need to browse around.

David Jones

Attention is fast but volition is slow

A random scan is a quicker way to find items in a display than a systematic search.

How swiftly can the object of your attention be changed? Consider two ways to deploy attention: it can be commanded from place to place by a deliberate act of will, or it can run freely without specific instruction. Here we use a visual search task to show that deliberate movement of attention is significantly slower because of an internal limit on the speed of volitional commands.

Extensive research on visual search tasks such as finding the letter F in Fig. 1a has revealed that stimuli of this sort can be searched at a rate equivalent to one letter every 25–50 ms (ref. 1). The order of search, even of eye movements, is influenced by stimulus salience and eccentricity, but is otherwise random through the set of salient loci². Why is there sparse evidence for systematic scanning of search displays? We argue that searching is free-running ('anarchic') because commanded, ordered deployment of attention is so much slower than anarchic deployment that it is faster overall to make many anarchic attentional deployments than fewer orderly ones.

In our 'command' condition, observers knew in advance where and when their attention needed to go. The 12 observers viewed each of 8 frames like those shown in Fig. 1a–d for 53 ms, followed by a mask which varied in duration from trial to trial. One target, a letter Y or N, was present on each trial. This target could appear only at position 1 (12 o'clock) on frame 1, position 2 on frame 2, and so on. If an observer deployed attention at the correct rate around the circle, attention would be on the target when it appeared (position 4 in Fig. 1d), otherwise the task was impossible. Observers were trained to move attention clockwise in time to a spatially uninformative tone. A 'staircase' procedure was used to measure the maximum speed that permitted 70% correct performance; software constraints limited the maximum rate to 80 ms per frame.

For comparison, observers were tested in a 'random anarchy' condition. Here, the Y or N was present on all frames but moved among three locations (different on each trial) between frames. Attention freely sampled the display at random. Randomization thwarted any parallel accumulation of information and prevented observers from attending to a single location and waiting for the target to arrive³. We asked observers to fixate; the letters were large enough to read without fixation and, critically, the stimulus configuration was the same in the 'random anarchy' and 'command' condi-

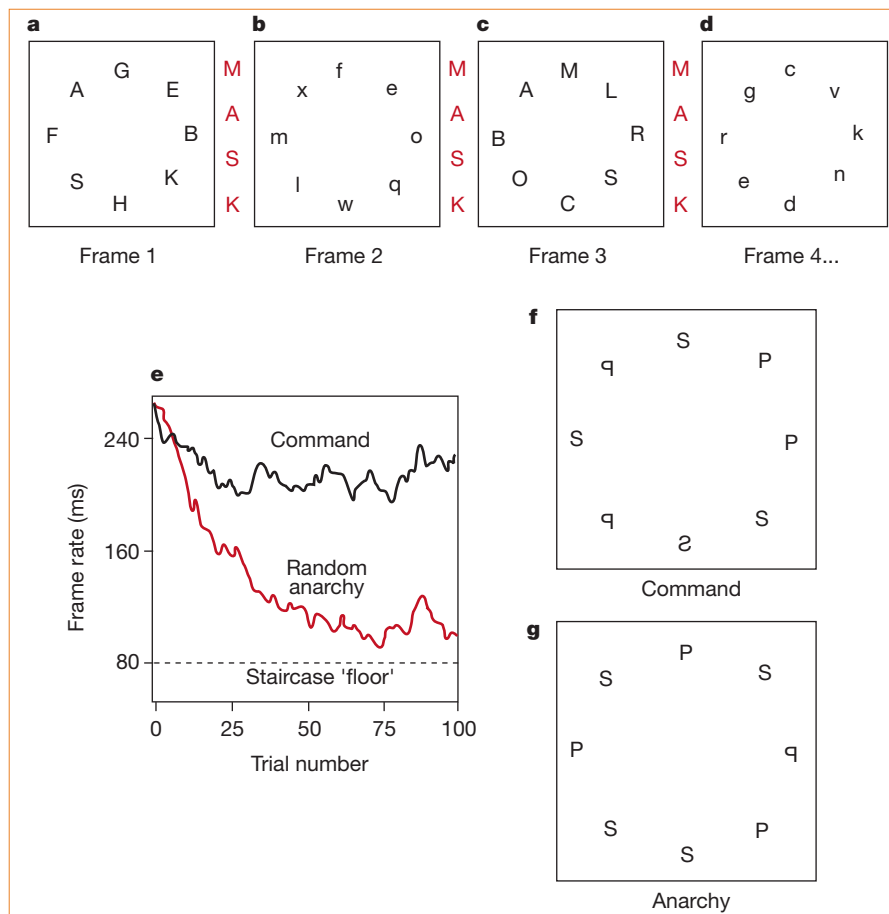


Figure 1 'Commanded' search is slower than 'anarchic' search. **a–d**, Four of eight frames in a 'command' condition trial. A 'Y' or 'N' can appear only at position *N* on frame *N* (position 1 corresponds to 12 o'clock). Here the target is in frame 4. Masks separate the frames. 'Random anarchy' uses the same stimuli, with a Y or N present in a random location on each frame. **e**, Staircase test results show that 'command' is much slower than anarchic deployment of attention. **f**, A static 'command' paradigm: search clockwise from the top to identify the first mirror-reversed letter. **g**, Static 'anarchy': find the only mirror-reversed letter. Again, commanded deployments are much slower than anarchic ones.

tions. Any eye movement or other perceptual limitations in the 'command' condition would also apply to 'random anarchy'.

A control experiment with static letters confirmed that these stimuli are searched at a standard rate of 41 ms per item. The average minimum 'commanded' rate was 217 ms per frame, or 217 ms per item. 'Random anarchy' produced a 105 ms per frame rate, or 105–187 ms per item, depending on assumptions about visual search (Fig. 1e). This is conservative because of the 80-ms-per-frame limit. Other 'anarchy' conditions without this constraint yielded estimates of between 13 and 44 ms per item. Three other 'command' conditions yielded maximum rates of 206, 345 and 195 ms per item (Fig. 1f, g shows an alternative method).

Our measures of 'commanded' attention are therefore slower by up to an order of

magnitude than any of our measures of anarchic attention. Our results indicate that visual attention may be deployed quickly and automatically by the salience of stimuli. Deliberate, volitional shifts of attention, however, can only be executed much more slowly, explaining why observers instinctively search unsystematically. Anarchy is faster than order in this case.

Jeremy M. Wolfe*†, **George A. Alvarez†**, **Todd S. Horowitz***†

*Harvard Medical School, †Center for Ophthalmic Research, Brigham and Women's Hospital, 221 Longwood Avenue, Boston, Massachusetts 02115, USA
e-mail: wolfe@search.bwh.harvard.edu

1. Wolfe, J. M. in *Attention* (ed. Pashler, H.) 13–74 (Psychology Press, Hove, UK, 1998).
2. Noton, D. & Stark, L. in *Perception: Mechanisms and Models—Readings from Scientific American* (eds Held, R. & Richards, W.) 218–227 (Freeman, San Francisco, 1971).
3. Horowitz, T. S. & Wolfe, J. M. *Nature* **394**, 575–577 (1998).

Biogeography

A marine Wallace's line?

As most coral reef organisms with a pelagic larval phase are presumed to be readily dispersed between distant populations, sea-surface current patterns should be crucial for predicting ecological and genetic connections among threatened reef populations¹. Here we investigate this idea by examining variations in the genetic structuring of populations of the mantis shrimp *Haptosquilla pulchella* taken from 11 reef systems in Indonesia, in which a series of 36 protected areas² are presumed to be connected by strong ocean currents. Our results reveal instead that there is a strong regional genetic differentiation that mirrors the separation of ocean basins during the Pleistocene low-sea-level stands, indicating that ecological connections are rare across distances as short as 300–400 km and that biogeographic history also influences contemporary connectivity between reef ecosystems.

Strong currents in Indonesian waters (Fig. 1a; up to 1 m s^{-1})³ should facilitate the dispersal of marine larvae and are thought to be responsible for the invisibility of Wallace's line, the sharp biogeographic break that separates terrestrial forms from eastern and western Indonesia, in marine communities. Drifters have traversed 1,500 km through the Celebes Sea and Makassar Strait in four weeks⁴, indicating that planktonic larvae may travel great distances, yielding high connectivity between distant populations. However, fish^{5,6} and invertebrate⁷ larvae in the Caribbean and Australia show a surprisingly small amount of movement, challenging the idea of a strict association between dispersal potential and realized movement between marine populations. We therefore measured the genetic structuring of populations of the stomatopod *H. pulchella*, a benthic reef crustacean with a 4–6-week planktonic larval period and a dispersal potential that is conservatively estimated at 600 km.

Surprisingly, our results indicate that there are sharp genetic breaks among oceanographic regions. Populations north and south of the Flores and Java Seas, as close as 300 km apart, are distinguished by a broad genetic break perpendicular to Wallace's line ($F_{ST}=0.821$ in analysis of molecular variance, $P<0.001$; ref. 8) (Fig. 1b). Additionally, northern populations are significantly differentiated ($F_{ST}=0.184$, $P<0.001$), with genetically distinct populations (Fig. 1b) in the Bay of Tomini and in the Celebes, Flores and South China Seas, contradicting the hypothesis of strong connectivity among localities.

The geographic distribution of geneti-

cally distinct populations reflects the greater isolation of ocean basins during Pleistocene low-sea-level stands⁹. The 'northern' and 'southern' clades may be relics of Indian and Pacific Ocean populations separated by the emergence of the Sunda and Sahul continental shelves during glacial maxima (Fig. 1c). Populations are also genetically distinct in the basins of the Celebes Sea, Tomini Bay and the South China Sea, which were partially enclosed during low-sea-level stands and have long been thought to be crucibles of species formation^{10,11}.

If the observed genetic differentiation arose during periods of low sea level, it is surprising that dispersal following 6,000–10,000 years of modern oceanographic conditions has not erased these historical boundaries. Genetic breaks have been found between Indian and Pacific Ocean populations that are separated by thousands of kilometres¹². However, the sharp genetic break described here that separates 'northern' and 'southern' populations a mere 300 km apart indicates the presence of a localized biogeographic boundary, suggesting the presence of a marine equivalent of Wallace's line. Our results indicate that predictions about connectivity based on generalized ocean

currents should also consider the biogeographic and oceanographic history.

Not all *H. pulchella* populations are genetically distinct, however. In the Spermonde archipelago (D in Fig. 1a), a collection of over 160 low coral islands in southwest Sulawesi, populations are genetically homogeneous over 10–100 km (5 populations, $n=76$, $F_{ST}=-0.03$, $P=0.76$). These populations and the Pantaloan population (B in Fig. 1a) 400 km upstream are dominated by a single mitochondrial DNA haplotype (54% and 50%, respectively) that is the molecular ancestor of most northern haplotypes (Fig. 1b). Although dominant in the Spermonde and Pantaloan populations, this haplotype is also present at low frequency in populations from the South China Sea, the Celebes Sea and the Lesser Sunda Islands. Likewise, a haplotype common in Riau (G in Fig. 1a) and Belitung (E, F) occurs at low frequency in the Spermonde and Talaud (A) (Fig. 1c). The distribution of these haplotypes indicates that gene flow is possible between distant populations, even if only rarely.

Information from other taxa should reveal whether these findings apply to other Indonesian reef species, but the association of stomatopod populations with old ocean

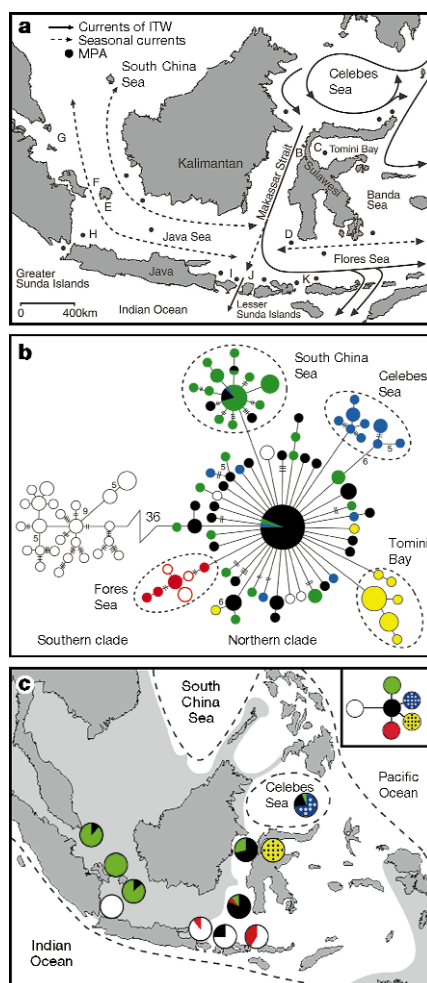


Figure 1 Ocean currents and genetic structure of stomatopod populations. **a**, Dominant currents (solid lines) and seasonal currents (dashed lines) of the Indonesian throughflow (ITW)³. Filled circles, marine protected areas (MPA)². Northern populations sampled were from regions marked A ($n=20$), B ($n=10$), C ($n=21$), D (76 individuals from 5 different reef areas), E ($n=10$), F ($n=10$) and G ($n=21$); southern populations were from regions H ($n=14$), I ($n=20$), J ($n=5$) and K ($n=6$). **b**, Unrooted minimum-spanning network (constructed using the program *Min-spn*; L. Excoffier) depicting the genetic relationship between 106 unique mitochondrial haplotypes from 213 individuals sampled from 11 regions throughout Indonesia. Haplotypes were identified by using the polymerase chain reaction to amplify 625 base pairs of the gene encoding cytochrome oxidase I (ref. 14) followed by automated sequencing on an ABI 377. The sizes of open circles (southern populations) and filled circles (northern populations) are proportional to haplotype frequencies and are separated by one mutational step, unless otherwise indicated by hatch marks or numerals. Dashed lines highlight clades found exclusively or primarily in one geographic region. Circles: red, Flores Sea; yellow, Tomini Bay; filled black, Makassar Strait; green, South China Sea; violet, Celebes Sea; white, Sunda Islands. **c**, Map of Indonesia at 18,000–20,000 years before present, when sea levels dropped by 130 m to expose the Sunda and Sahul continental shelves⁹ (light grey). Dashed lines highlight distinct ocean basins formed during the lowering of sea levels. Pie diagrams indicate relative haplotype frequencies. Inset, simplification of **b**.

basins suggests that reef populations throughout Indonesia cannot simply be assumed to be interconnected units; marine reserves need to be designed^{1,13} that also take biogeography and historical oceanography into account.

Paul H. Barber*, Stephen R. Palumbi*, Mark V. Erdmann†, M. Kasim Moosa‡

*Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA

e-mail: pbarber@oeb.harvard.edu

†Department of Integrative Biology, University of California, Berkeley, California 94720, USA

‡Pusat Penelitian dan Pengembangan Oseanologi, PO Box 4801 JKTF, Jakarta 1048, Indonesia

1. Roberts, C. M. *Science* **278**, 1454–1457 (1997).
2. Llewellyn, G., Jepson, P., Schweithelm, J. & Kirtchener, D. *Conservation Assessment of Wallacea Bioregion* (Wallacea Program Report, 1999).
3. Wyrki, K. *NAGA Report, Volume 2: Physical Oceanography of the South East Asian Waters* (Scripps Institute of Oceanography, San Diego, 1961).
4. Lukas, R. et al. *J. Geophys. Res.* **96**, 7089–7104 (1991).
5. Swearer, S. E., Caselle, J. E., Lea, D. W. & Warner, R. R. *Nature* **402**, 799–802 (1999).
6. Jones, G. P., Milicich, M. J., Emslie, M. J. & Lunow, C. *Nature* **402**, 802–804 (1999).
7. Knowlton, N. & Keller, B. D. *Bull. Mar. Sci.* **39**, 213–223 (1986).
8. Excoffier, L., Smouse, P. E. & Quattro, J. M. *Genetics* **4**, 479–491 (1992).
9. Porter, S. C. *Quat. Res.* **32**, 245–261 (1989).
10. McManus, J. W. *Proc. 5th Int. Coral Reef Symp.* **4**, 133–138 (1985).
11. Randall, J. E. *Zool. Stud.* **37**, 227–268 (1998).
12. Palumbi, S. R. *Coral Reefs* **16** (suppl.), 47–52 (1997).
13. Palumbi, S. R. in *Marine Community Ecology* (eds Berness, M., Gaines, S. & Hay, M.) (Sinauer, Sunderland, MA, in the press).
14. Barber, P. H. & Erdmann, M. V. *J. Crust. Biol.* **20**, 20–36 (2000).

Tokaimura accident

Neutron dose estimates from 5-yen coins

After a criticality accident at a nuclear-fuel processing facility in Tokaimura, Japan, at 10:35 Japanese standard time on 30 September 1999, neutron emission continued until early the next morning. To estimate the neutron dose to the surrounding population, we collected 5-yen coins (37 weight per cent zinc) from places likely to have been exposed to radiation as a result of the accident and determined the amount of ⁶⁵Zn generated in them in the nuclear reaction ⁶⁴Zn + n → ⁶⁵Zn by spectrometrically measuring the photons emitted by ⁶⁵Zn. Our results indicate that people evacuated to 350 m outside the irradiated area still received a significant neutron dose.

The Japanese 5-yen coin is about 22 mm in diameter and 1.5 mm thick, weighs 3.75 g and has a central hole 5 mm wide. We chose this coin for monitoring neutron exposure because it is widely circulated, the zinc content is precisely controlled, and the ⁶⁵Zn generated has a convenient half-life (244.1 days) and γ -ray energy emission

Table 1 ⁶⁵Zn in 5-yen coins from houses around the accident site

Distance (m)	Direction from facility	⁶⁵ Zn concentration (Bq per kg Zn)	Statistical error (%)
100 ± 10	SSW	39.6	6.4
100 ± 10	SSW	37.0	6.4
100 ± 10	SSW	39.4	1.6
120 ± 12	W	17.4	14.0
135 ± 10	SSW	16.4	36.0
170 ± 10	SSW	9.0	32.0
210 ± 12	SSW	5.4	45.0
230 ± 12	SSW	4.0	22.7
250 ± 20	SSE	4.0	9.6
255 ± 20	WNW	3.8	35.0
320 ± 15	WSW	2.0	10.5
320 ± 15	WSW	1.9	23.0
360 ± 15	SE	1.5	24.1
380 ± 6	SE	1.4	51.2
460 ± 20	N	1.1	24.0
550 ± 12	SSE	0.86	18.4
550 ± 12	SSE	0.76	17.9

*Samples that were separated and measured.

(1,115.5 keV). To obtain a record of the dosage of neutrons released as a result of the accident, we collected exposed coins from people's houses at distances 100–550 m from the facility.

Two types of detector were used to determine the concentration of ⁶⁵Zn in these coins: both are pure germanium detectors with a low background contribution, one of which (University of Tokyo) is covered with an inner copper shield 5 cm thick and a 5-cm layer of iron, with an outer lead casing of 10 cm — in this system, radon in the air is purged by using nitrogen gas; the other (Kyoto University) is covered with an inner copper shield of 10 cm and outer lead layer of 10 cm. Using these detectors, we tested 119 coins for a maximum of 1,000,000 seconds; some samples were measured in both systems to crosscheck for consistency. The scattering and self-absorption of 1,115.5-keV γ -rays in the coin samples were calculated by using the Monte Carlo method in three dimensions.

Table 1 shows ⁶⁵Zn concentrations measured in coins taken from houses in the neutron-affected areas around the nuclear-fuel processing facility. ⁶⁵Zn concentrations at distances ranging from 100 to 550 m from the plant dropped from 39.6 to 0.76 Bq per kg zinc, decreasing exponentially with increasing distance.

⁶⁵Zn concentrations and neutron dosage at various distances from the accident site have also been estimated by considering the neutron energy spectrum in relation to distance (T. Imanaka, personal communication); values were calculated with the neutron transport programs DOT and

MORSE (ref. 1). Using these data, we were able to estimate approximate neutron doses from the ⁶⁵Zn concentrations shown in Table 1.

Ambient dose equivalent values (in millisieverts at 1 cm depth) were calculated as about 220 mSv at 100 m, 6 mSv at 350 m and 1.8 mSv at 550 m. These estimates agree well with the early official estimates², which were revised to about half some weeks later³. Our estimated exposures indicate that those living 350 m away from the accident — the distance to which people were advised to evacuate — were also irradiated by neutron doses of over 1 mSv.

Our dose estimates include a bias because the neutron field in the environment might differ from that assumed in the calculation. If the neutron field is unimpeded, the ⁶⁵Zn in 5-yen coins can serve as an indicator of neutron dosage at various positions. Our results potentially offer information not only about the total neutron effect during the accident⁴ but also about shielding by modern Japanese houses, given that the coins were recovered from indoors.

Masuchika Kohno*, Yoshinobu Koizumi†

*Department of Nuclear Engineering, Kyoto University, Yoshida, Sakyo, Kyoto 606-8501, Japan

†Isotope Center, University of Tokyo, 2-11-16, Yayoi, Bunkyo, Tokyo 113-0032, Japan

1. Imanaka, T. *Gijutsu to Ningen* **28**, 56–65 (1999) (in Japanese).
2. *Dose Estimation in the Environment Neighboring with JCO, Tokai Branch after the Criticality Accident (Theoretical and Basic Data)* (report from the Science and Technology Agency of Japan, 4 November 1999) (in Japanese).
3. *Dose Estimation in the Environment Neighboring with JCO, Tokai Branch after the Criticality Accident (Theoretical and Basic Data)* (report from the Science and Technology Agency of Japan, 22 December 1999) (in Japanese).
4. *J. Envir. Radioactivity* **50**, 1–172 (2000).

brief communications is intended to provide a forum for short, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com

Atmospheric carbon dioxide concentrations over the past 60 million years

Paul N. Pearson* & Martin R. Palmer†

* Department of Earth Sciences, University of Bristol, Queens Road, Bristol BS8 1RJ, UK

† T. H. Huxley School, Imperial College, RSM Building, Prince Consort Road, London SW7 2BP, UK

Knowledge of the evolution of atmospheric carbon dioxide concentrations throughout the Earth's history is important for a reconstruction of the links between climate and radiative forcing of the Earth's surface temperatures. Although atmospheric carbon dioxide concentrations in the early Cenozoic era (about 60 Myr ago) are widely believed to have been higher than at present, there is disagreement regarding the exact carbon dioxide levels, the timing of the decline and the mechanisms that are most important for the control of CO₂ concentrations over geological timescales. Here we use the boron-isotope ratios of ancient planktonic foraminifer shells to estimate the pH of surface-layer sea water throughout the past 60 million years, which can be used to reconstruct atmospheric CO₂ concentrations. We estimate CO₂ concentrations of more than 2,000 p.p.m. for the late Palaeocene and earliest Eocene periods (from about 60 to 52 Myr ago), and find an erratic decline between 55 and 40 Myr ago that may have been caused by reduced CO₂ outgassing from ocean ridges, volcanoes and metamorphic belts and increased carbon burial. Since the early Miocene (about 24 Myr ago), atmospheric CO₂ concentrations appear to have remained below 500 p.p.m. and were more stable than before, although transient intervals of CO₂ reduction may have occurred during periods of rapid cooling approximately 15 and 3 Myr ago.

More than a century has passed since Arrhenius¹ proposed that early Cenozoic warmth was caused by high levels of carbon dioxide in the atmosphere and Chamberlin² suggested a variety of geological processes that could affect atmospheric carbon dioxide concentrations (p_{CO_2}). However, there is still disagreement as to the most significant controls on p_{CO_2} over long timescales. Some authors have stressed the importance of changing inputs to the atmosphere such as volcanic and hydrothermal outgassing^{3,4} or metamorphic decarbonation reactions⁵, while others have focused on outputs such as the weathering of silicate minerals and limestone formation^{6–8} or organic carbon burial^{9–11}. The level of early Cenozoic p_{CO_2} and when it might have declined is also still under debate^{5–7,10,11}.

The boron-isotope ($\delta^{11}\text{B}$) approach to p_{CO_2} estimation relies on the fact that a rise in the atmospheric concentration will mean that more CO₂ is dissolved in the surface ocean, causing a reduction in its pH. We are able to estimate the pH of ancient sea water by measuring the boron-isotope composition of calcium carbonate ($\delta^{11}\text{B}_{\text{cc}}$) precipitated from it. This is because boron in aqueous solution occurs as two species, $\text{B}(\text{OH})_3$ and $\text{B}(\text{OH})_4^-$, between which the equilibrium is strongly pH-dependent over the natural acidity range of sea water. Furthermore, there is a pronounced isotopic fractionation between the species of approximately -19.5‰ , so that the $\delta^{11}\text{B}$ of each species is highly dependent¹² on pH. Because boron incorporation into marine carbonates is predominantly from $\text{B}(\text{OH})_4^-$, $\delta^{11}\text{B}_{\text{cc}}$ is a sensitive pH indicator^{13,14}. The pH of sea water is governed by the carbonate equilibria, such that for a given pH value it is possible to calculate the aqueous CO₂ concentration and thereby make quantitative estimates^{15–20} of atmospheric p_{CO_2} .

The pH and aqueous CO₂ concentration of the surface ocean vary spatially because of factors such as deep-water upwelling, local productivity regimes and freshwater inflows. To arrive at pH estimates that most closely reflect atmospheric p_{CO_2} , it is necessary to measure the $\delta^{11}\text{B}$ of carbonates that were precipitated far from coastal influences and sources of upwelling. The ideal setting is in the low-latitude gyre systems, where a mixed layer of warm, low-density sea water in contact with the atmosphere generally overlies colder deep waters with little intermixing. Such environments

support abundant planktonic foraminifera (a group of microscopic protists) that secrete calcite (CaCO_3) shells. The shells fall to the sea floor, from which a record of upper-ocean pH of many millions of years can be obtained.

We analysed the $\delta^{11}\text{B}$ of monospecific sample splits of surface mixed-layer dwelling foraminifera from 32 sediment samples from the open tropical Pacific, spanning the past 60 Myr, augmenting data from six other previously studied samples^{19,20} (Table 1). It is necessary to use various species because none survived the entire time interval. Each monospecific sample consists of more than 100 individual shells that calcified at different times of the day and year over a period of several thousand years, so the small deviations from equilibrium pH and aqueous CO₂ that constantly occur in the upper ocean are likely to be averaged. Analytical methods were the same as previously published¹⁹ and boron-isotope ratios are quoted relative to standard NBS SRM 951.

The study sites (Ocean Drilling Program Sites 865, 871 and 872) are all from the sedimentary caps of flat-topped seamounts in the tropical north Pacific gyre, and are characterized by exceptionally good carbonate preservation. They have been the subject of considerable previous geochemical work, including isotopic, trace element and microstructural characterization of the foraminifer shells^{21–25}. There is no evidence from this work that diagenetic alteration has affected shell chemistry. We do not know of any similarly well preserved carbonates from the open ocean in the age range of 25–33 Myr ago.

In order for a pH value to be calculated from a $\delta^{11}\text{B}_{\text{cc}}$ measurement, it is necessary to assume a value for the boron-isotope composition of sea water ($\delta^{11}\text{B}_{\text{sw}}$); currently this value is $+39.5\text{‰}$, but it may have changed through time. Fortunately, geologically rapid fluctuations in $\delta^{11}\text{B}_{\text{sw}}$ are unlikely to have occurred as dissolved boron has a long residence time²⁶ in the ocean of approximately 20 Myr. We can estimate past $\delta^{11}\text{B}_{\text{sw}}$ values by utilizing the fact that different planktonic foraminifera species calcify from the surface mixed layer to the low pH conditions below the thermocline. The magnitude of the pH decline is controlled mainly by the local level of biological production, because

more productive locations have a greater flux of sinking organic particles, which oxidize in the water column, reducing pH at depth. Although there is good evidence that the sites we analysed were always oligotrophic (with a maximum apparent oxygen utilization between 50 and 150 $\mu\text{M kg}^{-1}$), in six previously studied time windows in the Cenozoic we have observed substantial differences in the $\delta^{11}\text{B}_{\text{cc}}$ differential between surface and deep planktonic environments^{19,20}. As we have argued²⁰, such differences can be explained as having been caused by variation in $\delta^{11}\text{B}_{\text{sw}}$. Hence, we use the value of the $\delta^{11}\text{B}_{\text{cc}}$ differential to model $\delta^{11}\text{B}_{\text{sw}}$ for each time window and interpolate $\delta^{11}\text{B}_{\text{sw}}$ values for the other samples (Table 1).

Changes in sea surface temperature (SST) also affect pH and p_{CO_2} estimates because the boric acid and carbonate equilibria are temperature dependent. However, oxygen isotopic studies have shown that the tropical SST has remained surprisingly constant through the Cenozoic despite pronounced cooling in the high latitudes²⁷. We have assumed a SST of 27 °C for all our samples.

The $\delta^{11}\text{B}$ of surface-dwelling foraminifera may not provide a good estimate of pH in the mixed layer if the microenvironment of the foraminifer cell has a distinctive pH because of photosynthesis by algal symbionts. A symbiont “vital effect” on $\delta^{11}\text{B}_{\text{cc}}$ has been suggested for the analogous but much larger aragonitic skeletons of corals²⁸. Microsensor studies of the living symbiotic species *Globigerinoides trilobus* (also known as *G. sacculifer*) and related taxa show that pH increases over short distances as the foraminifer shell is approached^{29,30}. However, there is evidence that *G. trilobus* precipitates its shell with no boron-isotope vital effect^{16–18}. Furthermore, because our analyses of various extinct species produces an oceanographically consistent pH–depth profile²⁰, we infer that Palaeogene surface-dwelling species of the genera *Acarinina* and

Morozovella also precipitated their shells at or close to boron-isotopic equilibrium.

However, one of the species studied, upper Eocene *Hantkenina alabamensis*, yielded extremely negative $\delta^{11}\text{B}_{\text{cc}}$ values in all three samples analysed that cannot be interpreted using the pH model (see Table 1). As there is no textural or geochemical reason to suspect diagenetic alteration²⁵, we conclude that a strong vital effect fractionation of the boron isotopes occurred in this species. One possibility is that *H. alabamensis* (which has a very involute shell and large enveloping final chambers) resorbed earlier-formed calcite late in its life cycle, thereby reworking isotopically light boron. These data are excluded from subsequent geochemical calculations.

Figure 1 illustrates the record of sea surface pH that we have reconstructed from the $\delta^{11}\text{B}_{\text{cc}}$ data. Our finding that the pH was significantly lower in the early Palaeogene than in the Neogene period has implications for a wide range of biological and chemical processes in the ocean. For example, the speciation of a number of dissolved elements in sea water (such as P, N, Mn and Se) undergoes variations over the range of pH values (7.4–8.3) calculated in this study. The surface ocean would not have been so acidic as to inhibit biological calcification, but other more subtle effects on plankton ecology and sedimentation patterns may have occurred.

Reconstructing seawater carbon chemistry

The pH of the surface ocean is one of four key variables that define its carbonate chemistry, the others being the concentration of aqueous CO_2 , the amount of total dissolved inorganic carbon (ΣCO_2) and the total alkalinity³¹. In order to calculate a value for p_{CO_2} it is necessary to assume a value for ΣCO_2 or alkalinity. Unfortunately, no reliable record of variations in either parameter

Table 1 Boron-isotope measurements of shallow-dwelling planktonic foraminifera and the calculated pH values

Age (Myr ago)	Top datum	Bottom datum	Hole	Sample (cm)	Species, size range (μm)	$\delta^{11}\text{B}_{\text{cc}}$	Analytical error (2 σ)	$\delta^{11}\text{B}_{\text{sw}}$ model	pH
0.085	0.06	0.11	871A	1H/1, 124–126	Various	24.9	0.4	39.10	8.10
0.98	0.47	1.37	871A	2H/2, 59–61	<i>G. trilobus</i> , 500–600	25.1	0.1	39.10	8.12
1.49	1.45	1.68	871A	2H/6, 59–61	<i>G. trilobus</i> , 500–600	25.2	0.3	39.10	8.13
3.00	2.9	3.1	871A	3H/2, 123–125	Various	25.9	0.4	39.10	8.21
3.31	3.1	3.5	872C	3H/2, 59–61	<i>G. trilobus</i> , 500–600	25.5	0.1	39.03	8.17
3.87	3.8	3.9	872C	3H/5, 118–120	<i>G. trilobus</i> , 500–600	25.1	0.2	38.91	8.14
6.00	5.8	7.1	871A	3H/5, 60–62	<i>G. trilobus</i> , 500–600	24.9	0.1	38.59	8.15
6.20	5.8	7.1	871A	3H/5, 123–125	Various	24.5	0.3	38.53	8.12
9.02	7.1	10.2	872C	5H/2, 14–16	<i>G. trilobus</i> , 500–600	25.0	0.1	38.30	8.20
10.39	10.2	10.4	872C	5H/6, 59–61	<i>G. trilobus</i> , 500–600	24.5	0.2	37.99	8.18
11.40	11.3	11.5	872C	6H/5, 20–22	<i>G. trilobus</i> , 500–600	24.4	0.4	37.79	8.19
11.81	11.5	12.6	871A	4H/5, 59–61	Various	24.1	0.3	37.70	8.16
13.06	12.6	13.1	871A	6H/6, 60–62	<i>G. trilobus</i> , 300–425	24.4	0.4	37.70	8.20
14.73	14.6	14.9	871A	7H/2, 124–126	<i>G. trilobus</i> , 300–425	25.5	0.4	37.70	8.31
14.96	14.9	15.2	871A	7H/5, 59–61	<i>G. trilobus</i> , 300–425	25.0	0.2	37.70	8.26
16.23	16.2	16.3	871A	8H/2, 59–63	Various	23.7	0.4	37.70	8.14
16.70	16.6	19.0	872C	11H/1, 20–22	<i>G. trilobus</i> , 300–425	24.3	0.3	37.75	8.18
18.38	16.6	19.0	872C	11H/6, 20–22	<i>G. trilobus</i> , 300–425	24.6	0.1	37.84	8.20
19.85	19.0	21.8	872C	12H/2, 78–80	<i>G. trilobus</i> , 300–425	24.7	0.4	37.97	8.20
21.70	19.0	21.8	872C	13H/3, 20–22	<i>G. trilobus</i> , 300–425	24.7	0.4	38.03	8.19
23.00	22.8	23.2	872C	13H/5, 20–22	<i>G. trilobus</i> , 300–425	24.3	0.1	38.12	8.12
23.51	23.2	23.7	872C	14H/4, 20–22	<i>G. trilobus</i> , 300–425	23.5	0.2	38.19	8.04
34.84	34.0	35.2	865C	3H/5, 110–112	<i>H. alabamensis</i> , >250	11.0*			
36.10	35.2	38.4	865C	4H/1, 110–112	<i>H. alabamensis</i> , >250	11.4*			
39.51	38.4	40.1	865C	4H/5, 110–112	<i>H. alabamensis</i> , >250	13.5*			
40.12	40.1	40.5	865C	5H/1, 110–112	<i>A. topilensis</i> , 300–425	23.0	0.5	39.24	7.80
42.52	40.5	43.6	865C	6H/2, 65–67	Various	25.0	0.5	39.40	8.07
44.26	43.6	45.8	865C	7H/1, 110–112	<i>A. topilensis</i> , 300–425	24.1	0.2	39.40	7.95
45.69	43.6	45.8	865C	7H/3, 110–112	<i>A. praetopilensis</i> , 300–425	23.1	0.6	39.40	7.79
46.07	45.8	46.1	865C	8H/3, 110–112	<i>A. praetopilensis</i> , 300–425	22.0	0.1	39.40	7.54
46.97	46.1	49.0	865C	8H/5, 110–112	<i>A. praetopilensis</i> , 300–425	24.4	0.4	39.40	7.99
50.33	49.0	50.4	865C	9H/5, 110–112	<i>A. praetopilensis</i> , 500–600	23.4	0.1	39.40	7.84
51.02	50.8	52.3	865C	10H/1, 110–112	<i>M. caucasica</i> , 500–600	23.9	0.5	39.40	7.92
52.22	50.8	52.3	865C	10H/5, 100–102	<i>M. caucasica</i> , 425–500	21.6	0.2	39.40	7.42
53.24	52.3	54.0	865C	11H/1, 110–112	<i>M. marginodentata</i> , 425–500	22.3	0.4	39.40	7.62
55.84	54.7	55.9	865C	12H/5, 110–112	<i>M. velascoensis</i> s.l., 425–500	21.8	0.3	39.40	7.48
57.12	57.1	57.5	865C	14H/3, 110–112	<i>M. velascoensis</i> s.l., 300–425	22.0	0.4	39.40	7.54
59.88	59.2	60.0	865C	15H/5, 110–112	<i>M. velascoensis</i> s.l., 300–425	21.6	0.2	39.40	7.42

The ages of each sample were calculated by linear interpolation between reliable biostratigraphical datums for Sites 871 and 872 (as determined by ref. 50) and for Site 865 (R. D. Norris & H. Nishi, unpublished data). Samples consisting of various species are the means of analyses presented in previous studies^{19,20}. Samples with an asterisk were excluded from subsequent geochemical calculations because they probably represent vital-effect fractionation of the boron isotopes. The $\delta^{11}\text{B}_{\text{sw}}$ value for each sample is calculated from the $\delta^{11}\text{B}_{\text{cc}}$ differential from surface to oxygen minimum zone at discrete levels, or interpolation between those levels following the method of ref. 20. *G.*, *Globigerinoides*; *H.*, *Hantkenina*; *A.*, *Acarinina*; *M.*, *Morozovella*.

exists over the timescale covered by this study. However, variation in the carbonate compensation depth (CCD) in the ocean has previously been inferred from historical patterns of carbonate sedimentation³², and this can be used to constrain alkalinity and ΣCO_2 .

Our approach was to model the ancient ocean by adjusting alkalinity so as to ensure that the water column is exactly saturated with respect to calcite at the lysocline (which we take to be 500 m shallower than the CCD). The depth of the lysocline depends on the calcium concentration of sea water, $[\text{Ca}]_{\text{sw}}$, as well as the carbonate equilibria. The most important process that may have caused changes in $[\text{Ca}]_{\text{sw}}$ and alkalinity over the timescales considered in this study is variation in the river water flux. For example, continental weathering was probably more intense during periods of warm climate and high p_{CO_2} , which would deliver more Ca and HCO_3^- to the ocean. Therefore we have assumed that $[\text{Ca}]_{\text{sw}}$ has remained proportional to alkalinity. We then calculate surface-ocean alkalinity and ΣCO_2 by assuming that increases in these parameters with depth have remained the same as the modern western equatorial Pacific Ocean, that is, $\Delta\Sigma\text{CO}_2$ from 0 to 4,000 m equals $340\text{ }\mu\text{M kg}^{-1}$, $\Delta(\text{alkalinity})$ from 0 to 4,000 m equals $75\text{ }\mu\text{eq kg}^{-1}$. Although this assumption may not be strictly correct, the relative homogeneity of sediment geochemistry in this area over the past 60 Myr (refs 21–24) means it has undergone considerably less variation in the factors that control depth variation in ΣCO_2 and alkalinity (principally productivity rates) than most other areas.

The calculated sea surface alkalinity record is shown in Fig. 2. The large amplitude variation in alkalinity over relatively short time periods in the Palaeogene is implied by the combined observations that the CCD was constant³² but surface ocean pH varied markedly. However, we note that the CCD record for the Palaeogene Pacific Ocean is relatively poorly constrained, and it is possible that an updated study would imply less rapid alkalinity variation.

Having established values for surface ocean pH and alkalinity, it is possible to calculate aqueous CO_2 and atmospheric p_{CO_2} . The p_{CO_2} record is shown in Fig. 3a, with an expansion for the Neogene in Fig. 3b. We emphasize that the error bars in Figs 1 to 3 reflect only the analytical uncertainties in determining $\delta^{11}\text{B}_{\text{cc}}$. The absolute estimates must be treated with caution because there are other sources of error that are less easy to quantify. One important uncertainty is that the values of the various constants for carbonate equilibria in sea water have not been accurately measured at p_{CO_2} levels above 500 p.p.m. (ref. 33), and so are subject to revision. Other problems,

as discussed above, include taking correct values for SST, $\delta^{11}\text{B}_{\text{sw}}$, $[\text{Ca}]_{\text{sw}}$, local biological productivity and CCD, as well as accounting for possible species-specific vital-effect fractionation and diagenetic alteration of the boron-isotope ratios. Despite these difficulties, which are similar to those encountered in other lines of palaeoclimate research, we believe that the timing and direction of the changes in the properties illustrated in Figs 1 to 3 are robust and warrant discussion.

Causes of Cenozoic p_{CO_2} variation

Modelling of the atmosphere suggests that p_{CO_2} forces surface temperature in a logarithmic relationship. We use the equation of Kiehl and Dickinson³⁴ to estimate the level of greenhouse heating predicted by our p_{CO_2} estimates. We note that the relatively small changes in p_{CO_2} that occurred in the Neogene would have had a disproportionately large effect on surface temperature because absorption of outgoing radiation by CO_2 is further from saturation at low values. In Fig. 4, we compare this record with the oxygen-isotope ratio ($\delta^{18}\text{O}$) of deep sea benthic foraminifera^{35–37}, which is an important climate proxy that reflects both cooling and ice growth. Also indicated on Fig. 4 are some of the most important palaeoclimate events of the Cenozoic.

It has long been recognized that the early to early-middle Eocene (55 to 45 Myr ago) was a pivotal time in global climate evolution when a long-term cooling trend was initiated that ultimately resulted in the present glaciated Earth³⁸. There is abundant geological evidence for large reorganizations of the carbon cycle in this interval. Our data indicate that a substantial reduction in p_{CO_2} occurred at this time, but the decline was not continuous, and the relationship with deep-sea temperature and ice growth was not straightforward.

During the late Palaeocene and early Eocene there was large-scale volcanic activity associated with North Atlantic rifting³⁹. Enhanced volcanic outgassing of CO_2 may have been supplemented by magmatism and regional metamorphism in parts of the Himalayan belt¹¹ and North America⁵. Another source of CO_2 may have been the oxidation of methane released from storage either in wetlands⁴⁰ or from large seafloor gas hydrate reservoirs, as may have occurred in the short-lived Late Palaeocene thermal maximum event at 55 Myr ago⁴¹ and similar subsequent events in the early Eocene⁴². In contrast, there is much less geological evidence for high levels of CO_2 emission in the later middle and upper Eocene.

We note that the termination of North Atlantic volcanism at about 54–53 Myr ago⁴⁰ corresponds approximately to the initial drop that we record in p_{CO_2} (Fig. 2). However, despite growing evidence for a short phase of cooling in the earliest Eocene⁴³, diverse geochemical and palaeontological indices suggest that the peak of

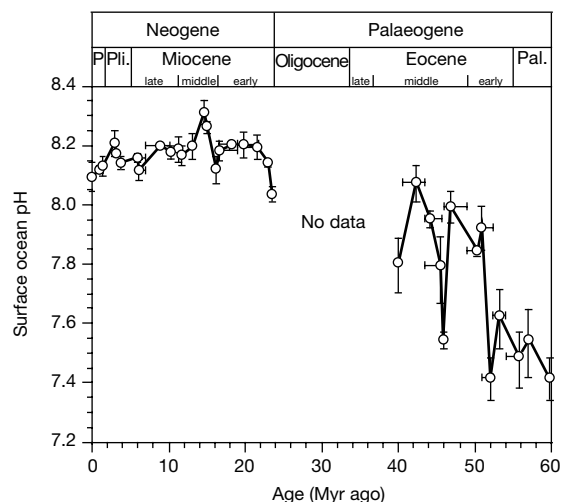


Figure 1 Sea surface pH for the past 60 Myr. Vertical error bars result from analytical error in determining $\delta^{11}\text{B}_{\text{cc}}$. Horizontal error bars represent the higher and lower biostratigraphical datums that constrain each sample (Table 1). Timescale according to ref. 49. P, Pleistocene; Pli., Pliocene; Pal., Palaeocene.

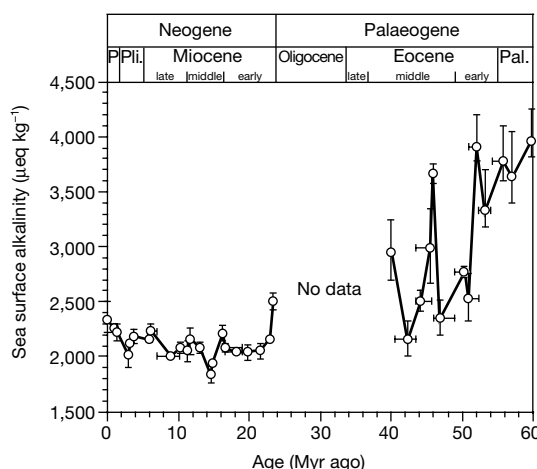


Figure 2 Sea surface alkalinity for the past 60 Myr. These values have been calculated by assuming that $[\text{Ca}]_{\text{sw}}$ has remained proportional to alkalinity. Epochs as for Fig. 1.

global warmth was not reached until about 54–50 Myr ago^{35–37}, a time for which we estimate relatively lower p_{CO_2} values (700–900 p.p.m.). This was a period of maximum warmth and continental humidity combined with intermittently low sea level, causing extensive mangrove swamp and coal formation¹⁰. Deep weathering on the exposed continents may have accelerated the drawdown of CO_2 . In addition, it has been proposed that continuing collision between India and Asia would have caused a progressive reduction in arc volcanism through this interval, combined with a switch from organic carbon (C_{org}) exhumation on the Indian continental shelf to C_{org} burial in large foreland basins, bringing about a reduction¹¹ in p_{CO_2} .

High-latitude cooling began at about the time of a sharp sea-level regression at 50 Myr ago that corresponds to a widespread hiatus in marine sediments^{35–37}. This event is the first of four major steps³⁶ (as highlighted on Fig. 4) that punctuate the long-term history of Cenozoic cooling. McGowan¹⁰ has proposed that a reverse greenhouse effect operated at that time, caused by high levels of siliceous plankton productivity in the oceans and correspondingly enhanced

rates of marine C_{org} burial. Our data do not support a precise covariation of p_{CO_2} and temperature; indeed we record a p_{CO_2} peak during the cooling phase at approximately 45.5 Myr ago. Nevertheless a longer-term link between C_{org} burial and p_{CO_2} cannot be ruled out. We note that a major transition in Cenozoic $\delta^{34}\text{S}_{\text{sw}}$ occurred 55–45 Myr ago as would result from extended deposition of high levels of pyrite in anoxic marine sediments⁴⁴, and this interval corresponds to a general decline in p_{CO_2} .

Overall, the p_{CO_2} values for the early Cenozoic show considerably more variability than do the values for the late Cenozoic. This may be partly due to the magnification of errors that occurs at low pH values, but it may also reflect greater instability of the global carbon system during warm periods. The relative constancy of p_{CO_2} values in the later Cenozoic is circumstantial evidence that p_{CO_2} is stabilized by a homeostatic feedback mechanism involving the greenhouse effect⁴⁵. This is because radiative forcing is amplified at low p_{CO_2} , hence small variations would be damped more efficiently by homeostasis.

Our oldest samples from this set indicate that there was a

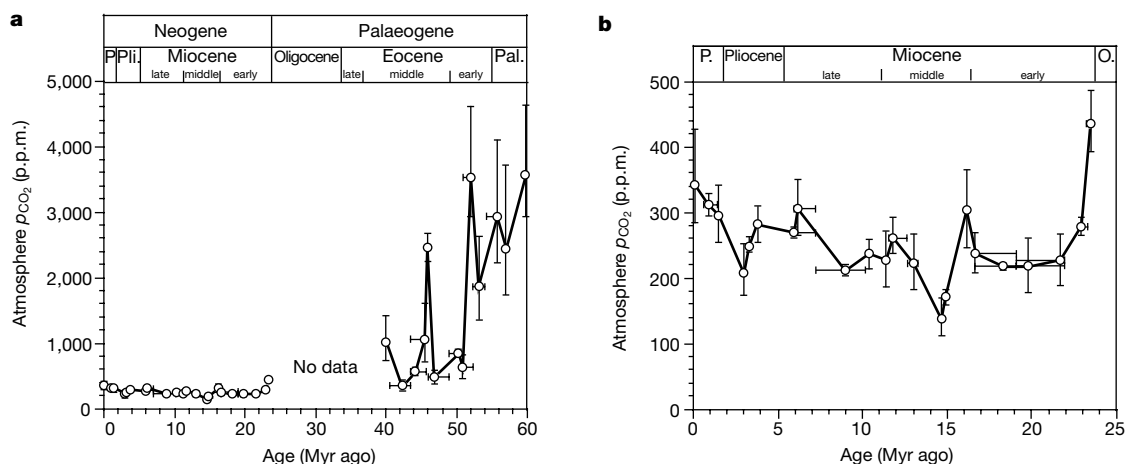


Figure 3 Record of atmospheric carbon dioxide for the past 60 Myr. **a**, The whole record; **b**, expansion for the past 25 Myr. We note that the nature of the $\text{pH}-\delta^{11}\text{B}_{\text{cc}}$ relationship means that analytical errors result in much larger uncertainties in the calculated p_{CO_2} at

lower pH values. These values were calculated using a modified version of the carbonate equilibria presented in ref. 33. Epochs as for Fig. 1; O., Oligocene.

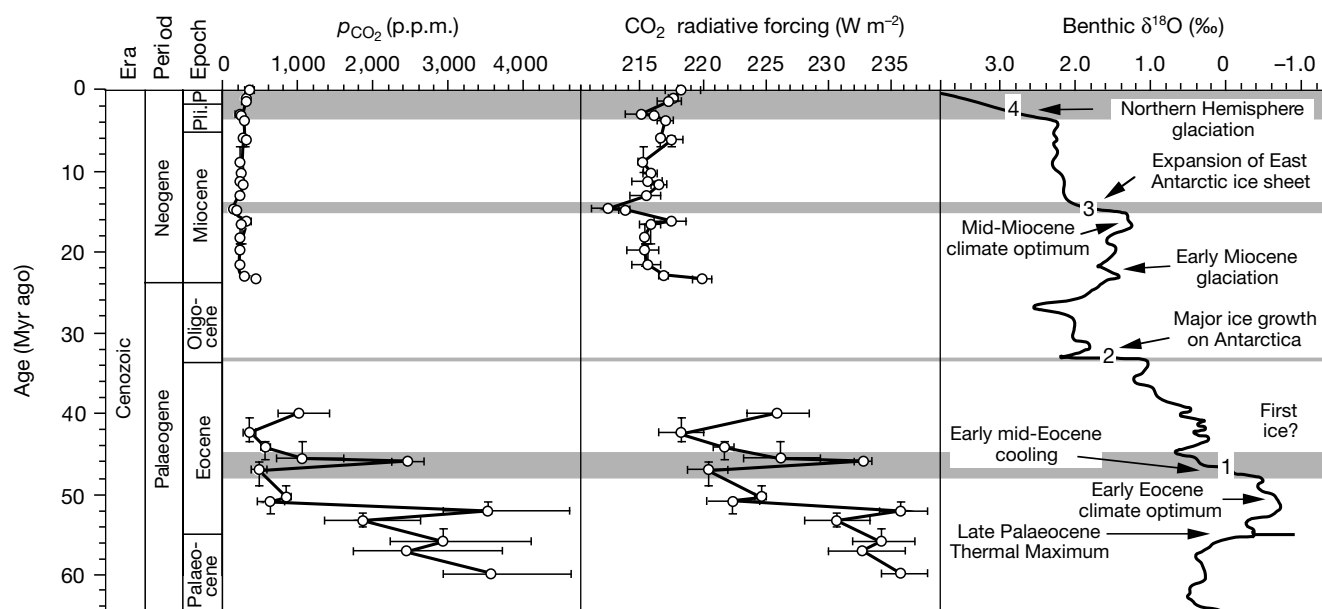


Figure 4 Carbon dioxide levels and Cenozoic climate change. The p_{CO_2} record is converted to radiative forcing (a measure of global warming) using the equation of Kiehl and Dickinson³⁴. The benthic foraminifer $\delta^{18}\text{O}$ time series is a smoothed record of many

observations summarized in refs 35–37. The trend towards more positive $\delta^{18}\text{O}$ results from a combination of deep-sea cooling and global ice volume increases. Four major steps (numbered 1 to 4) in $\delta^{18}\text{O}$ are indicated.

reduction in p_{CO_2} in the lowermost Miocene. This was a time of high-latitude cooling that may have been associated with enhanced glaciation on Antarctica³⁶. However, the most dramatic cooling phase in the Miocene was at 16–14 Myr ago, which has been interpreted as a major expansion of the East Antarctic Ice Sheet and increased formation of Antarctic Bottom Water⁴⁶. This intensification of bottom-water formation would have been accompanied by accelerated upwelling of nutrient-rich water, which we would expect to have boosted productivity and caused a temporary reduction in atmospheric p_{CO_2} . A major increase in C_{org} burial in diatomites around the Pacific rim has been recorded at this time⁹. Our $\delta^{11}B_{cc}$ data suggest a transient p_{CO_2} decline from about 300 p.p.m. to 140 p.p.m. However, there is little support in our data (or the similar p_{CO_2} record produced recently⁴⁷ on the basis of biomarker carbon isotopes) for a large and permanent drop in p_{CO_2} during the middle Miocene, as has previously been suggested^{8,9,15}.

Another prominent step in the global cooling trend occurred in the late Pliocene (between about 4 and 2 Myr ago)³⁶ which culminated in the onset of major glaciation in the Northern Hemisphere. Once again, this step was probably associated with enhanced deep-water formation (this time in the North Atlantic), which would have led to a transient increase in surface-ocean productivity and a lowering of p_{CO_2} . We note a downturn in our p_{CO_2} record at this time, from about 280 to 210 p.p.m. Denser sampling will obviously be needed to test all these possible linkages.

Change in the carbon dioxide concentration of the atmosphere is commonly regarded as a likely forcing mechanism on global climate over geological time because of its large and predictable effect on temperature⁴⁸. Our $\delta^{11}B_{cc}$ proxy for p_{CO_2} broadly confirms the prediction of Arrhenius¹ that early Cenozoic p_{CO_2} levels were often several times modern values, and that a strong greenhouse effect probably contributed to global warmth at that time. These 'super-greenhouse' conditions ($p_{CO_2} > 1,000$ p.p.m.) also imply considerably lower surface-ocean pH, higher alkalinity and higher levels of ΣCO_2 . We find that there was considerable fluctuation in these variables in the Palaeogene, but since the earliest Miocene the system seems to have been much more constant and more closely comparable to the present, despite continuing climate cooling. This suggests that other factors, such as complex feedbacks initiated by tectonic alteration of the ocean basins, were also important in determining global climate change. □

Received 2 November 1999; accepted 22 June 2000.

1. Arrhenius, S. On the influence of carbonic acid in the air upon the temperature on the ground. *Phil. Mag.* **41**, 237–279 (1896).
2. Chamberlin, T. C. An attempt to frame a working hypothesis of the cause of glacial periods on an atmospheric basis. *J. Geol.* **7**, 545–584 (1898).
3. Owen, R. M. & Rea, D. K. Sea floor hydrothermal activity links climate to tectonics—the Eocene carbon dioxide greenhouse. *Science* **227**, 166–169 (1985).
4. Berner, R. A., Lasaga, A. C. & Garrels, R. M. The carbonate-silicate geochemical cycle and its effect on atmospheric carbon dioxide over the past 100 million years. *Am. J. Sci.* **283**, 641–683 (1993).
5. Kerrick, D. M. & Caldeira, K. Metamorphic CO_2 degassing from orogenic belts. *Chem. Geol.* **145**, 213–232 (1998).
6. Brady, P. V. The effect of silicate weathering on global temperature and atmospheric CO_2 . *J. Geophys. Res.* **96**, 18101–18106 (1991).
7. Worsley, T. R., Moore, T. L., Fraticelli, C. M. & Scotese, C. R. Phanerozoic CO_2 levels and global temperatures inferred from changing paleogeography. *Geol. Soc. Am. (Special Paper)* **288**, 57–73 (1994).
8. Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122 (1992).
9. Berger, W. H. & Vincent, E. Deep-sea carbonates: reading the carbon-isotope signal. *Geologische Rundschau* **75**, 249–269 (1986).
10. McGowan, B. Silica burp in the Eocene ocean. *Geology* **17**, 857–860 (1989).
11. Beck, A., Sinha, A., Burbank, D. W., Seacombe, W. J. & Khan, S. in *Late Paleocene—Early Eocene Climatic and Biotic Events in the Marine and Terrestrial Records* (eds Aubry, M.-P., Lucas, S. G. & Berggren, W. A.) 103–117 (Columbia Univ. Press, New York, 1998).
12. Kikihana, H., Kotaka, M., Satoh, S., Nomura, M. & Okamoto, M. Fundamental studies on the ion exchange separation of boron isotopes. *Bull. Chem. Soc. Jpn* **50**, 158–163 (1977).
13. Hemming, N. G. & Hanson, G. N. Boron isotope composition and concentration in modern marine carbonates. *Geochim. Cosmochim. Acta* **56**, 537–543 (1992).
14. Hemming, N. G., Reeder, R. J. & Hanson, G. N. Mineral-fluid partitioning and isotopic fractionation of boron in synthetic calcium carbonate. *Geochim. Cosmochim. Acta* **59**, 371–379 (1995).
15. Spivack, A. J., You, C. F. & Smith, H. J. Foraminiferal boron isotope ratios as a proxy for surface ocean pH over the past 21 Myr. *Nature* **363**, 149–151 (1993).

16. Sanyal, A., Hemming, N. G., Hanson, G. N. & Broecker, W. S. Evidence for a higher pH in the glacial ocean from boron isotopes in foraminifera. *Nature* **373**, 234–236 (1995).
17. Sanyal, A. *et al.* Oceanic pH control on the boron isotopic composition of foraminifera: evidence from culture experiments. *Paleoceanography* **11**, 513–517 (1996).
18. Sanyal, A., Hemming, N. G., Broecker, W. S. & Hanson, G. N. Changes in pH in the eastern equatorial Pacific across Stage 5–6 boundary based on boron isotopes in foraminifera. *Glob. Biogeochem. Cycles* **11**, 125–133 (1997).
19. Palmer, M. R., Pearson, P. N. & Cobb, S. J. Reconstructing past ocean pH-depth profiles. *Science* **282**, 1468–1471 (1998).
20. Pearson, P. N. & Palmer, M. R. Middle Eocene seawater pH and atmospheric carbon dioxide concentrations. *Science* **284**, 1824–1826 (1999).
21. Bralower, T. J. *et al.* Late Paleocene to Eocene paleoceanography of the equatorial Pacific Ocean: Stable isotopes recorded at Ocean Drilling Program Site 865, Allison Guyot. *Paleoceanography* **10**, 841–865 (1995).
22. Pearson, P. N. & Shackleton, N. J. Neogene multispecies planktonic foraminifer stable isotope record, Site 871, Limalok Guyot. *Proc. ODP Sci. Res.* **144**, 401–410 (1995).
23. Israelson, C., Buchardt, B., Haggerty, J. A. & Pearson, P. N. Carbonate and pore water geochemistry of pelagic caps at Limalok and Lo-En guyots, western Pacific. *Proc. ODP Sci. Res.* **144**, 737–743 (1995).
24. Opdyke, B. N. & Pearson, P. N. Geochemical analysis of multiple planktonic foraminifer species at discrete time intervals. *Proc. ODP Sci. Res.* **144**, 993–995 (1995).
25. Coxall, H. K., Pearson, P. N., Shackleton, N. J. & Hall, M. A. Hantkeninid depth adaptation: an evolving life-strategy in a changing ocean. *Geology* **28**, 87–90 (2000).
26. Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell Scientific, Oxford, 1985).
27. Zachos, J. C., Lohmann, K. C., Walker, J. C. G. & Wise, S. W. Abrupt climate change and transient climates during the Paleogene: a marine perspective. *J. Geol.* **101**, 191–213 (1993).
28. Hemming, N. G., Guilderson, T. P. & Fairbanks, R. G. Seasonal variations in the boron isotopic composition of coral: A productivity signal? *Glob. Biogeochem. Cycles* **12**, 581–586 (1998).
29. Jorgensen, B. B. *et al.* Symbiotic photosynthesis in a planktonic foraminifer, *Globigerinoides sacculifer* (Brady) studied with microelectrodes. *Limnol. Oceanogr.* **30**, 1253–1267 (1985).
30. Rink, S. *et al.* Microsensor studies of photosynthesis and respiration in the symbiotic foraminifer *Orbulina universa*. *Mar. Biol.* **131**, 583–595 (1998).
31. Millero, F. J. Thermodynamics of the carbon dioxide system in the oceans. *Geochim. Cosmochim. Acta* **59**, 661–677 (1995).
32. Van Andel, T. J. Mesozoic–Cenozoic calcite compensation depth and the global distribution of calcareous sediments. *Earth Planet. Sci. Lett.* **26**, 187–194 (1975).
33. Wanninkhof, R., Lewis, E., Feely, R. A. & Millero, F. J. The optimal carbonate dissociation constants for determining surface water p_{CO_2} from alkalinity and total inorganic carbon. *Mar. Chem.* **65**, 291–301 (1999).
34. Kiehl, J. T. & Dickinson, R. E. A study of the radiative effects of enhanced atmospheric CO_2 and CH_4 on early Earth surface temperatures. *J. Geophys. Res.* **92**, 2991–2998 (1987).
35. Shackleton, N. J. Palaeogene stable isotope events. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **57**, 91–102 (1986).
36. Miller, K. G., Fairbanks, R. G. & Mountain, G. S. Tertiary oxygen isotope synthesis, sea level history, and continental margin erosion. *Paleoceanography* **2**, 1–19 (1987).
37. Zachos, J. C., Stott, L. D. & Lohmann, K. C. Evolution of early Cenozoic temperatures. *Paleoceanography* **9**, 353–387 (1994).
38. Frakes, L. A., Francis, J. E. & Syktus, J. I. *Climate Modes of the Phanerozoic* (Cambridge Univ. Press, Cambridge, 1992).
39. Ritchie, J. D. & Hitchen, K. in *Correlation of the Early Paleogene in Northwest Europe* (eds Knox, R. W. O., Corfield, R. M. & Dunay, R. E.) 63–78 (Special Publication 101, Geological Society, 1996).
40. Sloan, L. C. *et al.* Possible methane-induced polar warming in the early Eocene. *Nature* **357**, 320–322 (1992).
41. Dickens, G. R., O'Neill, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
42. Thomas, E., Zachos, J. C. & Bralower, T. S. in *Warm Climates in Earth History* (eds Huber, B. T., MacLeod, K. S. & Wing, S. C.) 132–160 (Cambridge Univ. Press, Cambridge, 2000).
43. Wing, S. L., Bao, H. & Koch, P. L. in *Warm Climates in Earth History* (eds Huber, B. T., MacLeod, K. S. & Wing, S. C.) 197–237 (Cambridge Univ. Press, Cambridge, 2000).
44. Paytan, A., Kastner, M., Campbell, D. & Thiemens, M. H. Sulfur isotopic composition of Cenozoic seawater sulfate. *Science* **282**, 1459–1462 (1998).
45. Broecker, W. S. & Sanyal, A. Does atmospheric CO_2 police the rate of chemical weathering? *Glob. Biogeochem. Cycles* **12**, 403–408 (1998).
46. Wright, J. D., Miller, K. G. & Fairbanks, R. G. Early and middle Miocene stable isotopes: Implications for deepwater circulation and climate. *Paleoceanography* **7**, 357–389 (1992).
47. Pagani, M., Arthur, M. A. & Freeman, K. H. Miocene evolution of atmospheric carbon dioxide. *Paleoceanography* **14**, 273–292 (1999).
48. Crowley, T. J. in *Warm Climates in Earth History* (eds Huber, B. T., MacLeod, K. S. & Wing, S. C.) 425–444 (Cambridge Univ. Press, Cambridge, 2000).
49. Berggren, W. A., Kent, D. V., Swisher, C. C. & Aubry, M.-P. A revised Cenozoic geochronology and chronostratigraphy. 129–212 (Special Publication 54, Society of Economic Paleontologists and Mineralogists, 1995).
50. Pearson, P. N. Planktonic foraminifer biostratigraphy and the development of pelagic caps on guyots in the Marshall Islands group. *Proc. ODP Sci. Res.* **144**, 21–59 (1995).

Acknowledgements

The authors contributed equally to this work. Samples were provided by the Ocean Drilling Program. We thank S. Cobb for assistance in sample preparation. This work was supported by the Natural Environment Research Council. P.N.P. is supported by a Royal Society University Research Fellowship.

Correspondence and requests for materials should be addressed to P.N.P. (e-mail: Paul.Pearson@Bristol.ac.uk).



Deep and stable interferometric nulling of broadband light with implications for observing planets around nearby stars

Kent Wallace, Graham Hardy & Eugene Serabyn

Jet Propulsion Laboratory, California Institute of Technology, MS 171-113, 4800 Oak Grove Drive, Pasadena, California 91109, USA

The number of indirectly detected planetary systems around nearby stars has grown tremendously since their initial discovery five years ago¹. But the direct observation of light reflected from these systems remains a formidable task, because of the high contrast ratios between them and their parent stars, and because of the tiny angular separations. Theoretically², these difficulties can be overcome by using a dual-aperture stellar interferometer in which the starlight is cancelled, or 'nulled', by broadband destructive interference, leaving the planet's light visible. Although the basic requirement of equal and oppositely directed electric fields is easy to state, an experimental demonstration of deep broadband nulling has been lacking, owing to difficulties engendered by the needs for extreme symmetry and stability, and low dispersion in the optical system. Here we report the deep (10^{-4}) and stable nulling of broadband (18% bandwidth) thermal light. These results validate the physical principles underlying future planet-searching interferometers, and our laboratory instrument will serve as a prototype for the nulling instrument to be implemented on the Keck interferometer in 2001.

The brightness contrast ratio of over $10^9:1$ between the Sun and the Earth at optical wavelengths implies that optically imaging such planets around nearby stars (separation is $0.1''$ at a distance of 10 pc) is impossible with present single-dish telescopes. On the other hand, the contrast ratio between a star and its attendant planets can theoretically be reduced² both by shifting to infrared wavelengths, where planetary thermal emission peaks, and by employing destructive interference between two collecting telescopes to cancel on-axis starlight. The problem then is to design a system in which the starlight cancellation is deep and broadband, as well as stable in time. Since the original conception³, several seemingly feasible approaches to deep, achromatic nulling have been proposed, but experimental verification has been lacking. These nulling schemes are based, respectively, upon a relative field-flip in a rotational shearing interferometer⁴⁻⁷ (RSI), a relative field inversion upon passage through focus⁸ and a phase retardation scheme^{9,10} in which each wavelength is delayed by a wavelength-independent π radians. Until now, the second¹¹ and third¹² approaches have achieved stellar rejections of approximately 5%, levels consistent with limitations set by optical quality and atmospheric fluctuations. Because mid-infrared null depths (defined as the inverse of the rejection ratio) of order 10^{-6} will eventually be needed for space-based terrestrial planet searches, it thus remains important to demonstrate that all of the experimental obstacles, such as fine path-length control, intensity balancing, low dispersion, wavefront clean-up, and polarization differences can indeed be overcome. Therefore we began with a series of laboratory experiments to validate the optical principles of deep nulling before the construction of an astronomical instrument. Optical wavelengths were used for these initial experiments because of their applicability to the planned nulling experiment on board NASA's Space Interferometer Mission¹³ (SIM), a location which will provide the first demonstration of nulling interferometry in space.

Our experimental set-up¹⁴ is based on the RSI approach^{4,5},

modified by the addition of polarization-compensating mirrors⁶ and the use of spatial filtering⁷ to effect wavefront clean-up. Our fibre-coupled RSI has already demonstrated very deep nulling of laser light^{14,15} with nulls over five orders of magnitude deep. The goals of our current set of experiments were to demonstrate that this nulling capability can be extended to a broad radiation bandwidth (that is, that an achromatic null can be achieved) and that the null can be stably maintained using only the weak signal from a thermal light source. In fact, we intentionally used a very faint white light signal, with a photon flux comparable to that expected from nearby stars on SIM. We also elected to begin with single-polarization light in order to test the feasibility of the optical scheme without the added complication of dual polarization operation (which would provide only a factor-of-two greater signal).

The details of our optical layout and the initial results with narrow-band laser radiation have been described previously^{14,15}, but several properties of the RSI warrant a quick review. First, the fundamental operational principle of the rotational shearing interferometer, the geometric field flip, is wavelength-independent. Second, since the beamsplitter is used in double-pass, each input beam is modulated by the same beamsplitter reflection/transmission product, thus automatically maintaining intensity balance for all wavelengths. Third, the reflections in the nuller are phase-compensated over all wavelengths by means of an additional fold mirror in each arm of the nuller⁷. Fourth, by coupling the focused output beam to a single mode fibre, only the core of the output point-spread function is detected, thus largely eliminating the effect of wavefront irregularities¹⁶. All of these special properties of the fibre-coupled RSI make it a singularly attractive architecture for nulling because it significantly relaxes the requirements on the optical coatings and surface quality. In terms of the experimental obstacles listed earlier, the only additional elements to consider are nanometre-level pathlength stability (to attain 10^{-4} nulls in the optical), and the minimization of dispersion.

It is worth noting the modifications made to the experiment in going from our previously reported laser results to wide-band operation. Our white-light source for these experiments is a quartz-halogen lamp coupled to the RSI through a single mode (for $\lambda > 500$ nm) optical fibre which provides a diffraction-limited star-like source. To eliminate dispersive dielectrics in the RSI, we accomplished the desired bandpass filtering and polarization selection before launching our light into free space. For this, we inserted

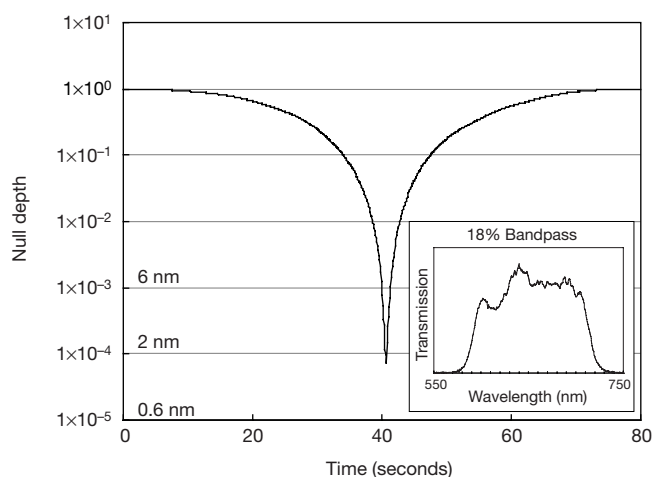


Figure 1 Intensity of the light transmitted through the nuller as the path length undergoes a linear sweep. The null depth is 7.4×10^{-5} in this scan. Each sample is 100 ms in duration. Inset, numbers on the y-axis represent the path-length match corresponding to the null depth scale; the chart is the spectral transmission of the light used in the measurements. The band-pass is roughly 590 nm to 710 nm, with a centre at 660 nm and a full-width at half-maximum of 120 nm, or 18%. The light is linearly polarized.

a miniature optical fibre bench with integral filters and polarizers between two fibres. We also added a fibre polarization-rotator to permit rotation of the single linear polarization state emerging from the fibre, thus allowing injection of white light with an arbitrary polarization orientation. We generally operate with s-plane polarization incident on the beamsplitter at 45° . The bandpass used in our experiment is roughly 'top hat' shaped, with a centre wavelength of 660 nm and a bandwidth of roughly 18% (Fig. 1). Our detector for the faint white light signal, a fibre-coupled avalanche photodiode (APD) was chosen to have sufficient dynamic range, without any gain change, to track the white light signal faithfully over at least five orders of magnitude (to see a 10^{-4} null unambiguously).

The location of the interferometer's zero optical path difference (OPD) fringe is found by translating one of the polarization compensation mirrors. A piezo-electric transducer (PZT) attached to a precision translation stage achieves a resolution of better than 1 nm. First, the deepest available destructive fringe is found by scanning in OPD through the white-light fringe packet. Dispersion in the system is determined by the asymmetry in the depth of the first destructive fringes to either side of the central fringe. Rotating the compensator to equalize the depths of these neighbouring fringes yields a local minimum in dispersion, and repeated optimization for successive fringes is used to find the global minimum null.

After precise alignment, a typical OPD scan through the null fringe is shown in Fig. 1. The null depth evident in this example is 7.1×10^{-5} , thus verifying the ability of the fibre-coupled RSI to provide an achromatic null over this wide bandpass. In comparison, the deepest destructive fringe that a conventional Michelson interferometer could generate for a bandwidth of 18% is 7×10^{-3} .

Path-length stabilization is the key to maintaining the null, and so to turning nulling into a useful technique for deep stellar rejection. Very small amplitude (~ 1 nm) path-length fluctuations between the two arms of the interferometer will degrade the null according to $N = (\varphi/2)^2$, where N is the null depth and φ is the phase error in radians. Because 1 nm is beyond the reach of typical metrology systems, the environment interior to the nuller must be made as stable as possible. To this end, the nuller is enclosed in Plexiglass and set upon a floating optical table with two levels of isolation (air legs and sorbothane pucks), the enclosure is lined with foam, and the optical mounts are all extremely rigid. Once the nuller is coarsely

aligned, fine alignment is done remotely from outside the sealed enclosure.

Even so, a path-length control scheme is needed to attain a stability level of the order of 1 nm. Several techniques have been proposed for this control, including use of a 'quadrature' output¹⁷ in an RSI, use of a waveband different for the nulling waveband¹², and of course, laser metrology. All of these techniques are under investigation, but for these initial experiments we opted for the much simpler approach of path-length dithering. This approach is standard in laser cavity stabilization, but as applied to nulling, the sign of the effect needs to be inverted to minimize the signal, and it must operate with far fewer photons. It might therefore be thought that the faint nulled signal is too weak to serve as an effective path-length error-sensing signal. However, since the error signal in dithering is given by the change in the photon flux relative to its value at the extremum, the error signals are exactly the same (except for sign) at a constructive peak and a destructive minimum. On the other hand, the photon noise is much reduced at null, resulting in an improvement in sensitivity (to OPD errors) that is proportional to the square root of the rejection ratio.

Figure 2 illustrates a dither-stabilized null in our fibre-coupled RSI. Under loop control, the null is stabilized to the 10^{-4} level (corresponding to a path-length stability of 2 nm), with an average null of approximately 7×10^{-5} . The average photon detection rate at null was 145 photons per second. As our dither frequency is 20 Hz, the path-length control loop thus operates successfully with only 7 photons per dither cycle. Thus nanometer-level stabilization of a white-light null is evidently possible with a minuscule number of photons.

We note that our flux levels at null are quite comparable to those expected from nearby stars on both SIM and the proposed Terrestrial Planet Finder (TPF)¹⁸ mission. Assuming an aperture size for SIM of 30 cm, and a system transmission of 10%, the photon detection rate from a G2 star located 10 pc away in a 20% wide band centred at $0.7 \mu\text{m}$ is 1.1×10^2 photons per second when nulling to the 10^{-4} level. Thus the performance level achieved in our experiments (including the level of the null, its stabilization, and the photon rate at null) already meets the stringent requirements of the nulling experiment for SIM (except for dual-polarization operation). For TPF, operating in the mid-infrared with larger (~ 3 m)

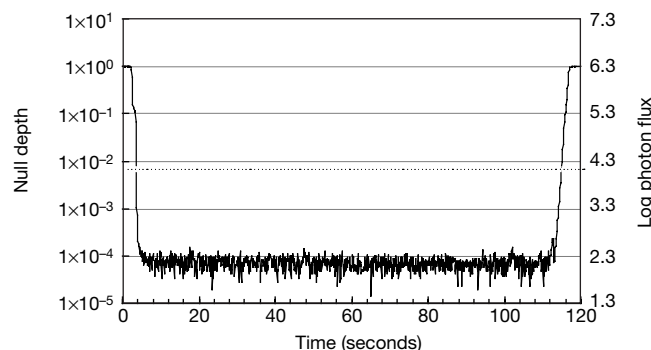


Figure 2 Illustration of the nuller output light under active control. The white-light signal begins high (at the first constructive fringe to one side of the null), is nulled (turned off) by changing the phase of the control loop by 180° , and then returned to its original level. Each sample is again 100 ms. Our analogue closed-loop dither scheme is implemented by applying a small-amplitude, sinusoidal voltage of frequency f to the piezo-electric transducer (PZT) that dithers the optical path difference (OPD) mirror with an amplitude of the order of 1 nm. The detected nuller optical output is then demodulated with a lock-in amplifier, which converts the component of the optical signal at frequency f to a d.c. error voltage. The OPD is driven to zero by adding this d.c. voltage to the dither signal applied to the PZT. The average null depth is about 7×10^{-5} under loop control with the worst null at about 1.4×10^{-4} and the best null of 1.4×10^{-5} . The horizontal line represents the

deepest null level possible with chromatic, destructive interference in a standard Michelson interferometer. The right-hand side gives the corresponding photon fluxes. A 10^{-4} null corresponds to a detected flux of 200 photons per second. The reported null depths do not account for two effects that degrade our measured nulling performance: detector non-linearity at high fluxes, and dark counts (counts due to noise in the absence of any photons) at low fluxes. In the first case, with increasing photon flux, the detector increasingly sees overlapping, indistinguishable arrivals leading to an undercount at the constructive peak of about 20% for our conditions. The counts at null reported here also do not correct for the intrinsic noise floor of the detector, which artificially raises the count level at null (by about 30%). Both of these effects degrade the measured null depth from the true ratio, so that our true nulls are thus probably deeper by about 50%.

apertures, the residual photon flux at null may be about an order of magnitude smaller due to the deeper (10^{-6}) nulls required in this case, but more sensitive path length control schemes are under development^{12,17}. So with one option verified, and several more under test, the stabilization of the null need no longer be perceived as a large difficulty.

Although the viability of an optical nulling experiment at the 10^{-4} null level as planned for SIM is thus established, a final question is how this work may extrapolate to other experiments. In the near term, the RSI approach will be applied in a mid-infrared nulling instrument for use with the Keck Interferometer on Mauna Kea¹⁹ in 2001. As we begin to move our experiment from visible to infrared wavelengths, we anticipate our task of achieving deep nulls will be made easier, owing to the improved optical wave front quality and the reduced effect of path length fluctuations at long wavelengths (nanometre-level path length control is sufficient for a 10^{-6} mid-infrared null if path length fluctuations dominate), but also made more difficult by the lack of components such as single mode fibres for the mid-infrared, and the need to operate cryogenically. However, except for dual-polarization operation, the technical uncertainties in the RSI approach have largely been removed.

The American and European space communities are considering planet finders based on infrared nulling interferometry for early in the next decade^{18,20}. The final architecture for such missions is far from decided, as is the nulling approach to be used. At this point, each of the proposed nulling techniques seems to have its own specific advantages and disadvantages. For example, the beamsplitter needs to be exactly 50/50 in the phase retardation approach, whereas this stringent requirement is absent in RSIs. On the other hand, RSIs yield two nulled outputs instead of one, and so are somewhat more complicated. Thus, although further development is certainly necessary to achieve the requisite broadband, dual-polarization, 10^{-6} nulls in the mid-infrared, our experiments do provide validation for most of the physical principles upon which the final nuller must be based: an achromatic phase-shift of π radians between the electric fields, broadband operation, active path-length control, dispersion minimization, and wavefront clean-up by spatial filtering. □

Received 24 March; accepted 28 June 2000.

1. Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
2. Bracewell, R. N. Detecting nonsolar planets by spinning infrared interferometer. *Nature* **274**, 780–781 (1978).
3. Bracewell, R. N. & Macphie, R. H. Searching for nonsolar planets. *Icarus* **38**, 136–147 (1979).
4. Diner, D. J. in *The Next Generation Space Telescope* (eds Bely, P. Y., Burrows, C. J. & Illingworth, G. D.) 133–141 (Space Telescope Science Institute, Baltimore, 1990).
5. Shao, M. in *The Next Generation Space Telescope* (eds Bely, P. Y., Burrows, C. J. & Illingworth, G. D.) 160–168 (Space Telescope Science Institute, Baltimore, 1990).
6. Diner, D. J., Tubbs, E. F., Gaiser, S. L. & Korechhoff, R. P. Infrared imaging of extrasolar planets. *J. Br. Interplanet. Soc.* **44**, 505–512 (1991).
7. Shao, M. in *Space Astronomical Telescopes and Instrumentation* (eds Bely, P. Y. & Breckinridge, J. B.) *Proc. SPIE* **1494**, 347–356 (1991).
8. Gay, J. & Rabbia, Y. Principe d'un coronographe interférentiel. *C.R. Acad. Sci. Paris* **322**, 265–271 (1996).
9. Angel, J. R. P. in *The Next Generation Space Telescope* (eds Bely, P. Y., Burrows, C. J. & Illingworth, G. D.) 81–94 (Space Telescope Science Institute, Baltimore, 1990).
10. Angel, J. R. P., Burge, J. H. & Woolf, N. J. Detection and spectroscopy of exo-planets like Earth. *Proc. SPIE* **2871**, 516–519 (1996).
11. Badouze, P. et al. First results with the achromatic interfero coronagraph. *Proc. SPIE* **3353**, 455–462 (1998).
12. Hinz, P. M. et al. Imaging circumstellar environments with a nulling interferometer. *Nature* **395**, 251–253 (1998).
13. Danner, R. & Unwin, S. (eds) *Space Interferometry Mission (SIM)* 103–124 (JPL Brochure 400–811, Jet Propulsion Laboratory, Pasadena, 1999).
14. Serabyn, E., Wallace, J. K., Hardy, G. J., Schmidlin, E. G. H. & Nguyen, H. Deep nulling of visible laser light. *Appl. Opt.* **38**, 7128–7132 (1999).
15. Serabyn, E., Wallace, J. K., Nguyen, H. T., Schmidlin, E. G. H. & Hardy, G. J. in *Working on the Fringe: Optical and IR Interferometry from Ground and Space* (eds Unwin, S. & Stachnick, R.) 437–442 (ASP Conf. Ser. 194, Astronomical Society of the Pacific, San Francisco, 1990).
16. Ollivier, M. & Mariotti, J.-M. Improvement in the rejection rate of a nulling interferometer by spatial filtering. *Appl. Opt.* **36**, 5340–5346 (1997).
17. Serabyn, E. Nanometre-level path-length control scheme for nulling interferometry. *Appl. Opt.* **38**, 4213–4216 (1999).
18. Beichman, A., Woolf, N. J. & Lindensmith, C. A. (eds) *Terrestrial Planet Finder (TPF)* (JPL Publication 99-3, Jet Propulsion Laboratory, Pasadena, 1999).

19. Colavita, M. M. et al. The Keck Interferometer. *Proc. SPIE* **3350**, 776–784 (1998).

20. Leger, A. et al. Could we search for primitive life on extrasolar planets in the near future? The DARWIN project. *Icarus* **123**, 249–255 (1996).

Acknowledgements

We thank the Keck Interferometer and SIM projects for their support of this work. We also thank H. Nguyen for help with data acquisition. The work reported here was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Correspondence and requests for materials should be addressed to K.W. (e-mail: James.K.Wallace@jpl.nasa.gov).

Superconductivity in molecular crystals induced by charge injection

J. H. Schön, Ch. Kloc & B. Batlogg

Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974-0636, USA

Progress in the field of superconductivity is often linked to the discovery of new classes of materials, with the layered copper oxides¹ being a particularly impressive example. The superconductors known today include a wide spectrum of materials, ranging in complexity from simple elemental metals, to alloys and binary compounds of metals, to multi-component compounds of metals and chalcogens or metalloids, doped fullerenes and organic charge-transfer salts. Here we present a new class of superconductors: insulating organic molecular crystals that are made metallic through charge injection. The first examples are pentacene, tetracene and anthracene, the last having the highest transition temperature, at 4 K. We anticipate that many other organic molecular crystals can also be made superconducting by this method, which will lead to surprising findings in the vast composition space of molecular crystals.

The experimental approach we followed is to inject a high concentration of charge carriers into organic molecular crystals of high quality, using a field-effect-transistor geometry (Fig. 1). There is a continuing effort to utilize a transverse static electric field in such an arrangement to modulate the properties of known superconductors². For example, the transition temperatures of copper oxide high-temperature superconductors can be varied significantly^{3–5}; this variation can be explained in terms of the known phase diagram of copper oxides, where the transition temperature changes systematically with the carrier concentration. More recently, the complete switching between an insulating and superconducting state has been reported in a fullerene field-effect transistor (FET)⁶. Our studies on organic molecular semiconductors, such as pentacene, revealed that low defect densities can be achieved in FET devices and that, in addition, the Fermi energy can be shifted through the bandgap leading to ambipolar transport⁷. Hence, the sign and concentration of charge carriers can be varied over an extremely wide range. The electronic structure of the

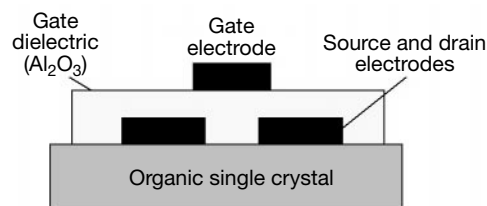


Figure 1 Schematic structure of an organic single crystal field-effect transistor.

polyacene crystals is characterized by an energy gap of several electron volts between the states derived from the HOMO (highest occupied molecular orbital) and the LUMO (lowest unoccupied molecular orbital): anthracene, 4.1 eV; tetracene, 3.1 eV; pentacene, 2.5 eV (ref. 8). Moreover, it has been suggested on theoretical grounds that superconductivity might occur in various molecular semiconductors due to electron–phonon coupling⁹. Here we report the observation of a superconducting state in an FET based on linear acenes, pentacene, tetracene and anthracene. The increase of the transition temperature from 2 to 4 K on going from pentacene to anthracene is in line with an electron–phonon coupling that increases as the conjugation length decreases and, therefore, supports the idea of superconductivity mediated by electron–phonon interactions.

The known organic superconductors can be classified as charge-transfer compounds, consisting of electronegative and electropositive entities^{10,11}. In contrast, the crystals investigated here are built of neutral organic molecules, and they are insulating when prepared in high purity. We have found that the charge transport in the acene crystals, such as naphthalene, anthracene, tetracene or pentacene, is dominated by electron–phonon interaction^{7,8,12,13}. Furthermore, it has been suggested that the infinite linear acene, polyacene, could also exhibit superconductivity^{14,15}. In our approach, we achieve charge transfer by an applied gate voltage in an FET geometry rather than by chemical interactions. This allows precise control of the charge per molecule without changing the molecular arrangement through dopants. Due to their good performances as FETs, anthracene, tetracene and pentacene⁷ were the materials of choice. We note in passing that the metallic layer produced in the channel of the device is in the closest possible contact with an insulating material (the insulating part of the same crystal); this offers the opportunity to study how the dielectric might influence superconducting pairing¹⁶.

We prepared single crystals of pentacene, tetracene and anthracene in a stream of hydrogen¹⁷. Typical crystals of these layered semiconductors exhibit surfaces of several mm², and are between 1 and 10 μm thick. Source and drain contacts were prepared on cleaved crystal surfaces (parallel to the molecular layers) by thermal evaporation through a shadow mask (channel length 25 μm , channel width 500–1,500 μm). A gate dielectric, Al₂O₃, was

sputtered onto the contacted surfaces. Finally, a gate electrode was deposited on top of the oxide (see Fig. 1). Measurements of these FETs were performed in the temperature range 1.7–300 K. In this geometry, the electrons injected into the FET channel move parallel to the layers of the crystals. The mobility within the layers is much higher than perpendicular to them. Ambipolar charge transport is obtained for pentacene⁷ as well as for tetracene and anthracene. Charge carrier mobilities increase from around 2 cm² V^{−1} s^{−1} at room temperature to several hundred cm² V^{−1} s^{−1} at low temperatures, typical for polyacene crystals^{7,12,13}. In the best crystals, we have measured electron mobilities up to 10⁴ cm² V^{−1} s^{−1} (ref. 18). The in-plane anisotropy of the charge carrier mobility is in the range of 50%. However, no significant influence of the in-plane anisotropy on the superconducting state has been observed. The carrier concentration in the FET channel can be adjusted by the gate bias, and concentrations in excess of 10¹⁴ electrons cm^{−2} can be induced. These densities correspond to about one electron per molecule, if we assume that only the topmost molecular layer takes part in the conduction¹⁹.

Figure 2 shows the channel resistance of a pentacene FET as function of temperature for two values of applied gate bias, 100 V and 120 V. The temperature dependence of the resistivity at these high carrier concentrations—near one electron per molecule ($> 10^{14}$ cm^{−2})—is typical of metallic transport. The residual resistivity ρ_0 for $T \rightarrow 0$ (several hundred $\mu\Omega$ cm) is 6–7% of the room-temperature value. Overall, the temperature dependence follows approximately a T^2 -law, yet a closer look (see Fig. 2 inset, where $(R(T) - R_0)$ is plotted against T) reveals that the resistivity varies more rapidly ($\propto T^{2.3}$) at high temperatures, and less rapidly ($\propto T^{1.8}$) at low temperatures. For a more detailed analysis, it will be necessary to correct for the effects of the pronounced thermal expansion of these crystals.

The significant finding of this study is shown in Fig. 2: for the higher gate voltage, the resistance below 2 K drops to zero within the experimental limit, suggesting a transition to a superconducting state. (These are two-point measurements by definition, and the experimental lower limit is approximately 0.01% of the normal-state resistance.) A similar transition is observed in tetracene and anthracene at a critical temperature T_c of 2.7 K and 4 K, respectively (Fig. 3). Figure 4 shows a detailed plot of the resistance as function of temperature and gate bias (or electron density per molecule). At a concentration of approximately one electron per molecule, the FET channel becomes superconducting below 2 K. After an abrupt onset, the transition temperature does not depend significantly on the

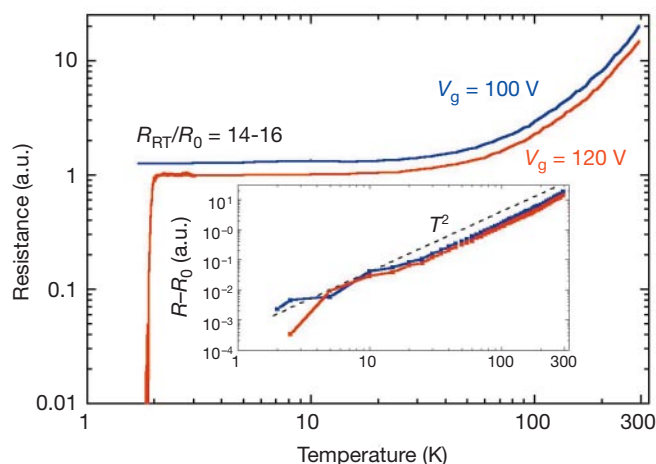


Figure 2 Channel resistance of a pentacene field-effect device at two gate voltages as a function of temperature. At these high gate voltages (V_g), close to one electron per molecule is injected and the material becomes metallic. The resistance decreases with decreasing temperature owing to the increasing electron mobility. For the higher carrier concentration (higher gate voltage) the resistance drops to zero below 2 K, indicating superconductivity. The inset shows $R - R_0$ (R_0 is $R(T \rightarrow 0)$) as a function of temperature. The dashed line corresponds to a T^2 -behaviour. R_{RT} , resistance at room temperature.

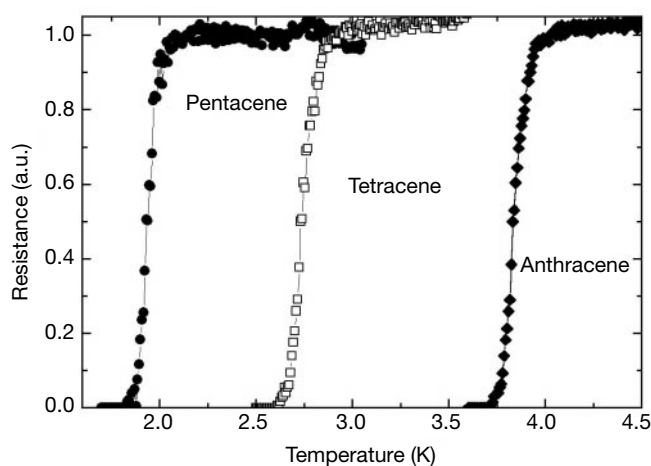


Figure 3 Channel resistance of pentacene, tetracene and anthracene field-effect devices as a function of temperature. The electron density in each channel is approximately one electron per molecule.

concentration. However, it was not possible to reliably achieve higher concentrations due to the breakdown of the dielectric. Using dielectric constants and high dielectrics with high breakdown fields, a more detailed investigation will be possible.

Figure 5 shows the dependence of the resistance on the magnetic field (perpendicular to the channel). The transition temperature is reduced, as is expected for a superconductor. The upper critical field H_{c2} can be deduced from the resistance data, resulting in an initial slope dH_{c2}/dT of approximately -3 T K^{-1} (see inset in Fig. 4). The standard extrapolation to zero temperature gives a value for $H_{c2}(0)$ of 4.1 T, corresponding to a coherence length of approximately 80 Å.

We can only speculate on the microscopic pairing mechanism, but it appears that the polarizability of the molecular crystal is a primary candidate. Assuming a conventional Bardeen–Cooper–Schrieffer mechanism with electron–phonon interaction, and the same density of states at the Fermi level, the increase of T_c from pentacene to anthracene can be ascribed to increasing electron–phonon coupling with decreasing conjugation length (or with number of delocalized π -electrons), in line with theoretical calculations^{9,14}. Further studies are needed, however, to shed more light on the microscopic aspects of pairing. One approach would be

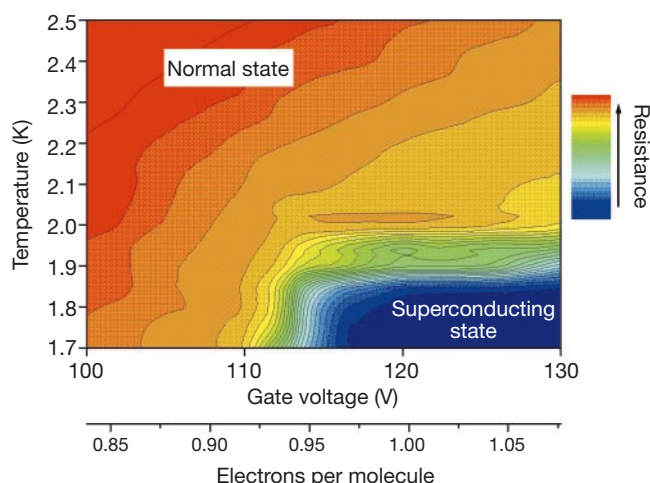


Figure 4 Channel resistance of a pentacene field-effect device as a function of temperature and gate bias. The transition to a superconducting state below 2 K and for gate biases above 112 V is clearly observable. The electron density per molecule is calculated assuming a homogeneous density only in the first molecular layer.

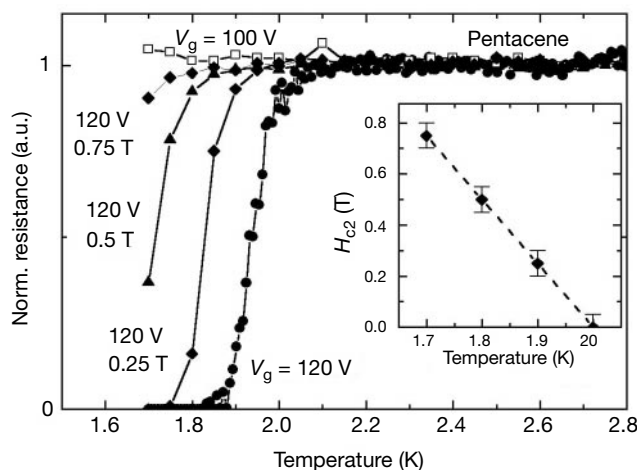


Figure 5 Channel resistance of a pentacene field-effect device as function of temperature and magnetic field (perpendicular to the channel). The inset shows the upper critical field as function of temperature (slope approximately -3 T K^{-1}), and a coherence length of 80 Å can be estimated for the superconductor.

to vary the carrier concentration over a wide range, and thus sweep the Fermi level through the density-of-states structure of the band. After this initial demonstration of superconductivity on relatively simple molecules, it would be worthwhile to pursue further this approach in order to explore the many possibilities offered by organic chemistry, using the powerful synthetic methods to engineer the electronic and vibrational properties of molecules. □

Received 18 April; accepted 17 July 2000.

1. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba–La–Cu–O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Glover, R. E. & Sherill, M. D. Changes in superconducting critical temperature produced by electrostatic charging. *Phys. Rev. Lett.* **5**, 248–250 (1960).
3. Mannhart, J., Ströbel, J., Bednorz, J. G. & Gerber, Ch. Large electric field effects in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ films containing weak links. *Appl. Phys. Lett.* **62**, 630–632 (1993).
4. Kossin, P. & Sorkin, B. Electric field effects in high T_c cuprates. *Phys. Rev. B* **58**, 5795–5802 (1998).
5. Ahn, C. H. *et al.* Electrostatic modulation of superconductivity in ultrathin $\text{GdBa}_2\text{Cu}_3\text{O}_{7-x}$ films. *Science* **284**, 1152–1155 (1999).
6. Schön, J. H., Kloc, Ch., Haddon, R. C. & Batlogg, B. A superconducting field-effect switch. *Science* **288**, 656–658 (2000).
7. Schön, J. H., Berg, S., Kloc, Ch. & Batlogg, B. Ambipolar pentacene field-effect transistors and inverters. *Science* **287**, 1022–1023 (2000).
8. Silinsh, E. A. & Capek, V. *Organic Molecular Crystals* (AIP, New York, 1994).
9. Devos, A. & Lannoo, M. Electron-phonon coupling for aromatic molecular crystals: possible consequences for their superconductivity. *Phys. Rev. B* **58**, 8236–8239 (1998).
10. Jerome, D., Mazaud, A., Ribault, M. & Beechard, K. Superconductivity in a synthetic organic conductor $(\text{TMTSF})_2\text{PF}_6$. *J. Phys. Lett.* **41**, L95–L98 (1980).
11. Ishiguro, T., Yamaji, K. & Saito, G. *Organic Superconductors* (Springer, Berlin, 1998).
12. Warta, W. & Karl, N. Hot holes in naphthalene: High, electric-field dependent mobilities. *Phys. Rev. B* **32**, 1172–1182 (1985).
13. Warta, W., Stehle, R. & Karl, N. Ultrapure, high mobility organic photoconductors. *Appl. Phys. A* **36**, 163–170 (1985).
14. Kivelson, S. & Chapman, O. L. Polyacene and a new class of quasi-one dimensional conductors. *Phys. Rev. B* **28**, 7236–7243 (1983).
15. Mishima, A. & Kimura, M. Superconductivity of the quasi-one-dimensional semiconductor polyacene. *Synth. Met.* **11**, 75–84 (1985).
16. Ginzburg, V. L. The problem of high temperature superconductivity. *Contemp. Phys.* **9**, 355–374 (1968).
17. Kloc, Ch., Simpkins, P. G., Siegrist, T. & Laudise, R. A. Physical vapor growth of centimeter-sized crystals of α -hexathiophene. *J. Cryst. Growth* **182**, 416–427 (1997).
18. Schön, J. H., Kloc, Ch. & Batlogg, B. Fractional quantum Hall effect in organic molecular semiconductors. *Science* **288**, 2338–2340 (2000).
19. Dodabalapur, A., Torsi, L. & Katz, H. E. Organic transistors: Two-dimensional transport and improved electrical characteristics. *Science* **268**, 270–271 (1995).

Acknowledgements

We thank E. Bucher for use of equipment, and H. Y. Hwang, D. W. Murphy, H. Störmer and C. M. Varma for discussions. J.H.S. was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to B.B. (e-mail: batlogg@bell-labs.com).

Room-temperature electronic phase transitions in the continuous phase diagrams of perovskite manganites

Young-Kook Yoo[†], Fred Duewer[†], Haitao Yang^{*}, Dong Yi^{*}, Jing-Wei Li^{*} & X.-D. Xiang^{*}

^{*} Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[†] Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA

Highly correlated electronic systems—such as transition-metal oxides that are doped Mott insulators—are complex systems which exhibit puzzling phenomena, including high-temperature superconductivity and colossal magnetoresistivity. Recent studies^{1–3} suggest that in such systems collective electronic phenom-

ena are important, arising from long-range Coulomb interactions and magnetic effects. The qualitative behaviour of these systems is strongly dependent on charge filling (the level of doping) and the lattice constant. Here we report a time-efficient and systematic experimental approach for studying the phase diagrams of condensed-matter systems. It involves the continuous mapping of the physical properties of epitaxial thin films of perovskite manganites (a class of doped Mott insulator) as their composition is varied. We discover evidence that suggests the presence of phase boundaries of electronic origin at room temperature.

There have been numerous studies aimed at understanding the relationship between charge dopant level n (or ionic radii)

and the physical properties of complex systems—particularly low-temperature thermodynamic phase transitions involving long-range order. Two such studies were conducted in copper oxides that showed high-temperature superconductivity (see, for example, ref. 4 and references therein) and colossal magnetoresistive manganites^{5–7}. These phase-diagram studies are very time consuming and deficient because they usually involve growing single-crystal samples of discrete (rather than continuous) compositions, and subsequently measuring physical properties at various temperatures. Our continuous phase diagram (CPD) method allows the continuous mapping of physical properties versus composition; this is much more time-efficient, thorough and systematic than the conventional approach.

The stability of the perovskite manganites against chemical substitutions provides an opportunity to study the effect of varying the parameters of a highly correlated electronic/magnetic system. Here we control the charge filling range n of Mott insulators by continuous charge doping over the entire range; we also control the hopping integral t through the substrate-induced anisotropic strain effect in epitaxial films and the average ionic radius of the A site. Our initial study uncovered several extremely narrow (in composition) phases with very different electronic properties. These narrow,

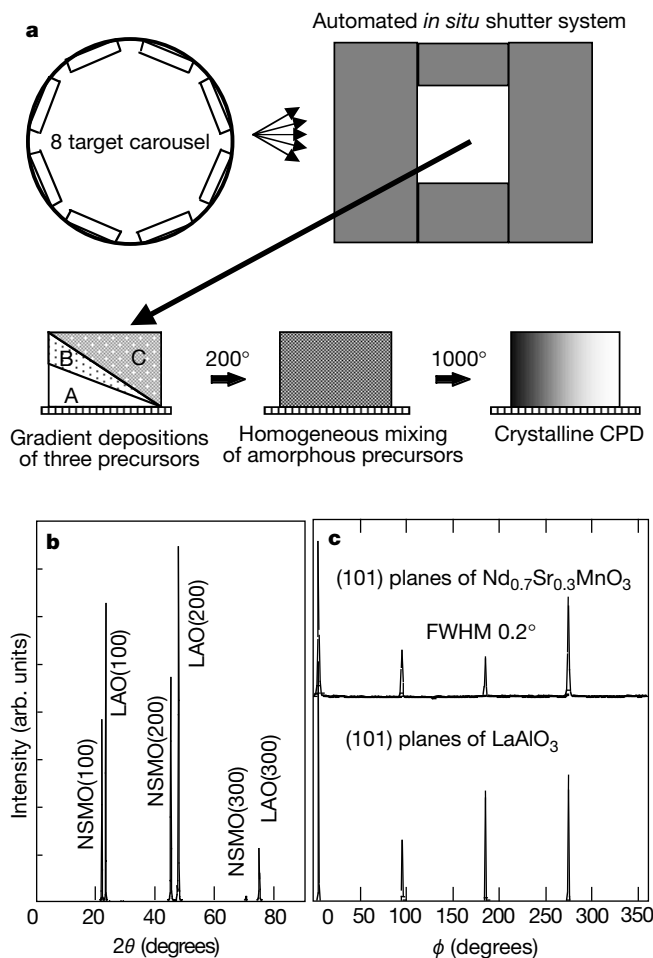


Figure 1 Multi-layer deposition and post-annealing synthesis of the CPD, with X-ray diffraction of a fabricated film. **a**, High-precision *in situ* shutter system in a pulsed laser deposition system. Eight-target carousel allows uninterrupted depositions without breaking the vacuum. Vertical shutters are used to define the width of each phase strip on a substrate while precise gradient profiles of three precursors are deposited with horizontal shutters moving across the phase strip at constant speed. The precursor films in this study were deposited at room temperature by a pulsed laser deposition system in a high vacuum ($\sim 10^{-7}$ torr). The forward-expanding plume in a high vacuum, coupled with scanning laser beam across the 2 inch $\times$ 2 inch targets during deposition, results in a deposition thickness uniformity of better than 1.5% over a 15 mm $\times$ 15 mm area, ensuring the accuracy in stoichiometry (easily controlled by shutter). Following the deposition, the sample was annealed at 200 °C for several days, then annealed at 400 °C for 30 h, followed by 2 h sintering at 1,000 °C. Low-temperature annealing is necessary to allow homogeneous mixing of precursors into an amorphous intermediate before crystallization at higher temperatures. **b**, The $\theta/2\theta$ XRD pattern and **c**, the ϕ -scan of the (101) plane of a $\text{Nd}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (NSMO) thin film made from three precursors of Nd_2O_3 , Mn_2O_3 and SrMnO_3 on (001) LaAlO_3 .

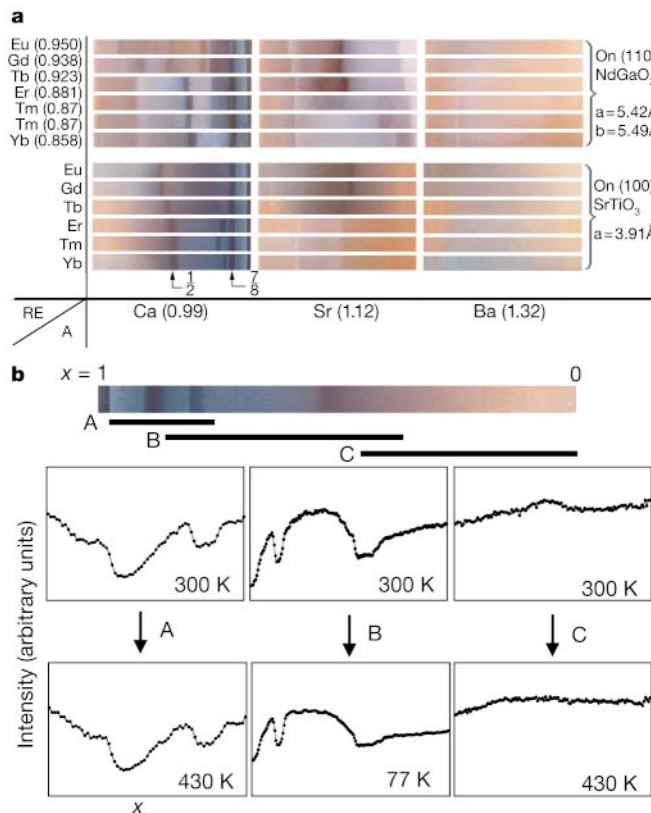


Figure 2 Photo-reflection images of $\text{RE}_{1-x}\text{A}_x\text{MnO}_3$ CPDs on NdGaO_3 and SrTiO_3 substrates, with the temperature-dependent study of $\text{Tm}_{1-x}\text{Ca}_x\text{MnO}_3$ CPD on SrTiO_3 . **a**, The room-temperature CCD colour photograph (photo-reflection image) of 36 CPDs on two different substrates under white light ((4.2–7.8) $\times 10^{14}$ Hz). The optical images were taken using a monochrome charge-coupled device (CCD) camera and a blue filter. Incident white light at 60° incidence was used to illuminate the sample and scattered light at 30° incidence was measured as a function of temperature from 77 to 430 K. The parentheses indicate the crystal ionic radii of the elements³⁰. The commensurate doping points of singular phases are denoted. The effective unit cell of (110) NdGaO_3 has $a' = b' = 3.86 \text{ \AA}$ ($a' = b' = \frac{\sqrt{a^2 + b^2}}{2}$). **b**, The room-temperature CCD colour photograph of $\text{Tm}_{1-x}\text{Ca}_x\text{MnO}_3$ CPD on SrTiO_3 under white light. Three doping regions are selected for blue light ((6.2–7.8) $\times 10^{14}$ Hz) photo-reflection at three different temperatures.

unexpected phases would be difficult to identify using conventional methods.

We determined CPDs of perovskite manganites ($\text{RE}_{1-x}\text{A}_x\text{MnO}_3$) in epitaxial thin-film form; here RE is Eu, Gd, Tb, Er, Tm and Yb, A is Ca, Sr and Ba, and x is varied continuously from 0 to 1 (Fig. 1a). Different phase diagrams were measured on six 15 mm by 15 mm substrates of two different single crystals, (100) SrTiO_3 , and (110) NdGaO_3 . As illustrated in Fig. 1a, gradients of RE oxides (Eu_2O_3 , Gd_2O_3 , Tb_4O_7 , Er_2O_3 , Tm_2O_3 and Yb_2O_3) are deposited as bottom-layer precursors. Each phase diagram was synthesized from gradient depositions of three precursors, using a high precision *in situ* linear shutter system (Fig. 1a) and subsequent *ex situ* post-annealing. Compared to the previously used composition-spread method by co-sputtering of multiple targets^{8–10}, our technique can easily generate precisely controlled stoichiometric profiles within a small area. This advantage is crucial when various single-crystal substrates need to be used (as in this study) for high-quality epitaxial film growth. Despite the fact that the crystalline compounds are formed *ex situ* rather than *in situ*, we found that appropriate post-annealing processes can generate high-quality epitaxial thin films.

Figure 1b shows a $\theta/2\theta$ X-ray diffraction (XRD) pattern for $\text{Nd}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ on an LaAlO_3 substrate which indicates that the film is (100) oriented (in order to study the crystalline quality of the samples, we made individual samples of various compositions selected from the phase diagrams under identical fabrication and processing conditions). Even when using a logarithmic scale intensity, we can hardly see any other impurity phase. Figure 1c shows ϕ -scans of the (101) planes, indicating that the film is in-plane aligned with

the substrate. We have confirmed similar epitaxial growth in many different compounds with different hole-doping levels.

To map the electronic properties of the CPDs, we choose two probes with very different energy scales (by 10^6): visible light and microwave. Visible colour photographs—optical-reflection images under white light, $(4.2\text{--}7.5) \times 10^{14}$ Hz—of CPDs were obtained as a function of temperature (Fig. 2). In order to map the electrical impedance at low frequencies, we used a scanning evanescent microwave probe operating at 2.2 GHz (refs 11, 12). Briefly, the probe design consists of a sharp metal tip protruding through the endwall of a coaxial resonant cavity. The high spatial frequency components of evanescent waves generated at the tip (spatial frequency is inversely proportional to tip radius) give rise to ‘super-resolution’ (exceeding the Abbe barrier), currently reaching $\lambda/10^5$ on dielectric samples. We note that λ is here the wavelength in the dielectric materials rather than the vacuum wavelength (‘super-resolution’ is usually not appropriate in metal samples, because the wavelengths in metals are less than $1\text{ }\mu\text{m}$). The propagating waves are confined within the cavity by a layer of thin metal shielding coated on the endwall. The instrument measures the complex impedance of the probe by monitoring the changes in resonant frequency (f_r) and quality factor (Q) of the cavity. The measured shift in f_r and Q can then be used to calculate the real and imaginary parts of the complex dielectric constant at f_r , ϵ_r and ϵ_i ($\epsilon_i = 4\pi\sigma/\omega$ for conducting samples, σ is the conductivity and $\omega = 2\pi f_r$) (Fig. 3b). The details of the instrument and quantitative analysis methods were published previously¹².

We present here detailed measurements of the room-temperature microwave impedance of a CPD of $\text{Er}_{1-x}\text{Ca}_x\text{MnO}_{3+\delta}$ (Fig. 3a). The

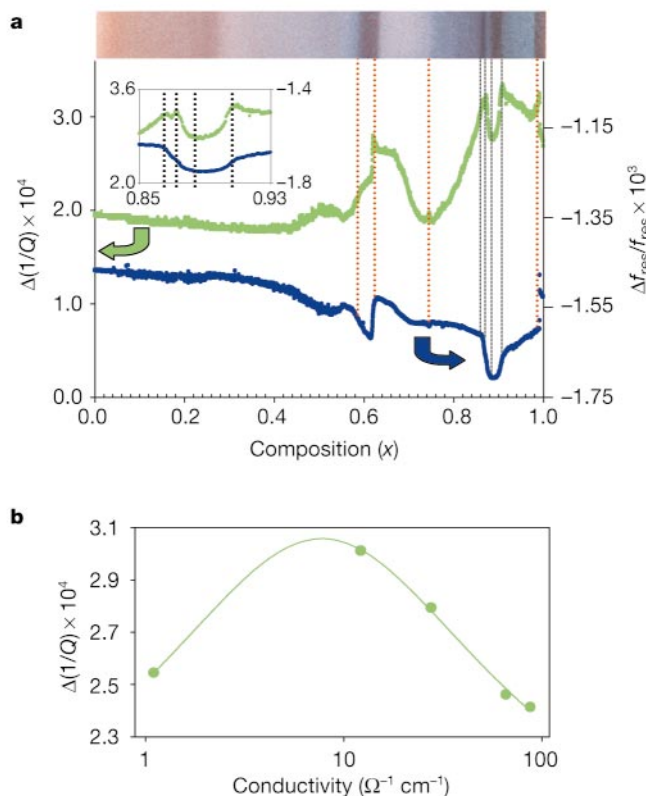


Figure 3 The microwave response of the $\text{Er}_{1-x}\text{Ca}_x\text{MnO}_3$ CPD on NdGaO_3 , and the calibration curve of microwave loss versus d.c. conductivity. The proximity of a sample to the tip changes the complex impedance of the probe, changing f_r and Q . In this study, the model analysis is based on a first-order approximation, which is valid when the conductivity is low. The shift in resonant frequency is a monotonic increasing function of the conductivity. The shift in quality factor is double-valued, approaching zero as the sample becomes insulating or highly conducting. The shift in quality factor has been

calibrated against the measured d.c. conductivity and is consistent with the theoretical analysis. **a**, The microwave response of the $\text{Er}_{1-x}\text{Ca}_x\text{MnO}_3$ CPD on NdGaO_3 . Microwave loss, $\Delta(1/Q)$, and frequency shift, $\Delta f_{\text{res}}/f_{\text{res}}$, are shown as a function of composition (x). The dashed lines are used to indicate phase boundaries. Inset, expanded-scale view of the curve near the commensurate singularity. **b**, Microwave loss, $\Delta(1/Q)$, versus d.c. conductivity ($\Omega^{-1}\text{ cm}^{-1}$). Filled circles show the measured microwave loss and the corresponding d.c. responses; the line is the theoretical fit.

room-temperature optical images of the complete sets of CPDs are provided for comparison (Fig. 2a). First, we observed many clear boundaries in both the optical and the microwave impedance images (dashed lines in Fig. 3a). Microwave loss peaks at these boundaries. X-ray data show no evidence of different structural phases beside perovskite throughout these CPDs. The existence of these boundaries suggests that complex electronic orderings (with fundamentally different physical properties) occur as a function of x .

Second, a very narrow insulating (or more accurately semi-conducting) strip within a highly conducting region (appearing as a greyish-blue strip within a black band in Fig. 2a), around 7/8 doping points of the CPDs of $\text{Tm}_{1-x}\text{Ca}_x\text{MnO}_3$ on NdGaO_3 substrate and $\text{Yb}_{1-x}\text{Ca}_x\text{MnO}_3$ on SrTiO_3 was observed at commensurate charge-filling point $x = 7/8$. The strip is so narrow in phase-width that it would have been very difficult to find it using the conventional practice of studying samples of discrete composition. We believe that this commensurate singularity point is related to the static charge (or spin) orderings observed in recent experimental and theoretical studies in both copper oxides^{1–3,13–20} and manganites^{21–25}. We expect to see a small structural anomaly associated with the singularity, owing to the strong coupling of charge ordering with the lattice at the commensurate point.

In Fig. 2b, different doping ranges are selected for temperature-dependent comparisons of blue light— $(6.2\text{--}7.8) \times 10^{14} \text{ Hz}$ —reflection images. In strip A, as the temperature is raised, there is clear evidence of broader phase boundaries (blurring of the phase), indicating the effect of thermal fluctuation. Increased temperature allows the coexistence of different phases over a broader range in composition. It also shows the high-energy scale of the electronic phases observed. In strip B, as the temperature is lowered, the suppression of intensity at the middle portion is clear. This tends to sharpen the phase boundary towards the right side (less doping) of the dark conducting phase. The conducting phase around $x = 1/2$ weakens at low temperature. In other manganite systems such as $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$, a similar phase near the 1/2 doping point was found to have two or more intrinsic competing phases (phase separation) with one phase dominant as the temperature is lowered. As the signal intensity is proportional to the amount of scattered light, the disappearance of this dip at low temperature indicates that a homogeneous phase dominates at 77 K. In strip C, the boundary (shown as a peak in intensity) disappears at high temperature.

The gradual change in ionic radii also induces unexpected abrupt changes in phase patterns. This effect can be observed in substitutions of both rare-earth and different divalent alkaline-earth elements. We see relatively smooth phase patterns and the absence of sharp transition boundaries in the Sr-doped system, and this trend is more pronounced in the Ba-doped system. The clear effect of the small t (or bandwidth), due to the smaller ionic size of Ca and the associated overlap of electron waves, is observed in the rich and diverse phase patterns in the Ca-doped system. The effect of substrate-induced stress can also be observed.

The doped Mott insulators are strongly correlated electronic systems, and the doped carriers tend to self-organize into highly anisotropic patterns. Upon doping into the parent Mott insulator compounds, the charge carriers tend to separate (at a large scale) into phases of different electronic natures, that is, phase separation. The long range Coulomb interaction tends to frustrate this phase separation and favour an anisotropic long-range order of charge or spin (stripe phase for example)¹³. More specifically, in manganese oxides the active orbital degree of freedom induces the highly anisotropic electron-transfer interaction, which gives rise to complex spin–orbital coupling effects (including stripe phases)^{26,27}. A possible model for the current observation of phase boundaries is the existence of different ground states of electronic self-organization due to corresponding orbital states mediated by long range Coulomb interaction. With the transfer interaction heavily depending on the doping level, various orbital ordered

and disordered states may exist with concomitant spin (and occasionally charge) orders. In $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ and $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$ CPDs, we have also observed both optical and electronic manifestation of such orbital states at room temperature with strong correlation to the low-temperature magnetic phases (ref. 28, and Y.-K.Y. *et al.*, manuscript in preparation).

However, the singular phases around $x = 7/8$ in the CPD of $\text{Er}_{1-x}\text{Ca}_x\text{MnO}_3$ on NdGaO_3 substrate and $\text{Yb}_{1-x}\text{Ca}_x\text{MnO}_3$ on SrTiO_3 substrate are intriguing (Fig. 2a). It seems that these singularities at 7/8 are due to the commensurate locking of orbital (charge) orderings to the lattice. If we accept this model, the highly conducting adjacent regions then suggest the existence of incommensurate, fluctuating orbital orderings. Therefore, the current observation provides preliminary evidence of the recently proposed¹ liquid-crystal phase (smectic phase) of orbital orderings in this doping range, widely believed to exist in copper oxides. The smeared-out critical divergence in microwave loss at the boundaries observed in this study is also consistent with this possibility. The thermodynamic nature of the transition is clearly shown by the temperature-dependent broadening or disappearance of the phase boundaries at higher temperatures. However, we believe that more studies are needed to confirm this suggestion.

The systematic experimental data provided here should help to identify new phenomena and elucidate the underlying physics of this complex system. The importance of the present work lies in the fact that we can continuously map the doping dependence of complex materials, instead of extrapolating from discrete doping points. For example, if doping-dependent electronic boundaries were also present in high- T_c copper oxides such as $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, the detailed study of the system at low temperature could be performed, and might reveal suspected quantum critical behaviour at the critical doping point of $x = 1/8$ (ref. 29). □

Received 17 January; accepted 27 May 2000.

- Kivelson, S. A., Fradkin, E. & Emery, V. J. Electronic liquid-crystal phases of a doped Mott insulator. *Nature* **393**, 550–553 (1998).
- Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
- Tranquada, J. M. *et al.* Coexistence of, and competition between, superconductivity and charge-stripe order in $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **78**, 338–341 (1997).
- Birgeneau, R. J. & Shirane, G. in *Physical Properties of High Temperature Superconductors* Vol. 1 (ed. Ginsberg, D. M.) 152 (World Scientific, Singapore, 1989).
- Schiffer, P., Ramirez, A. P., Bao, W. & Cheong, S.-W. Low temperature magnetoresistance and the magnetic phase diagram of $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$. *Phys. Rev. Lett.* **75**, 3336–3339 (1995).
- Urushibara, A. *et al.* Insulator-metal transition and giant magnetoresistance in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$. *Phys. Rev. B* **51**, 14103–14109 (1995).
- Tomioka, Y., Asamitsu, A., Kuwahara, H., Moritomo, Y. & Tokura, Y. Magnetic-field-induced metal-insulator phenomena in $\text{Pr}_{1-x}\text{Ca}_x\text{MnO}_3$ with controlled charge-ordering instability. *Phys. Rev. B* **53**, R1689–R1692 (1996).
- Sawatzky, E. & Kay, E. Cation deficiencies in r.f. sputtered gadolinium iron garnet films. *IBM J. Res. Dev.* **13**, 696–702 (1969).
- Hanak, J. J. The 'multiple-sample concept' in materials research: synthesis, compositional analysis and testing of entire multicomponent systems. *J. Mater. Sci.* **5**, 964–971 (1970).
- van Dover, R. B., Schneemeyer, L. F. & Fleming, R. M. Discovery of a useful thin-film dielectric using a composition-spread approach. *Nature* **392**, 162–164 (1998).
- Lu, Y. *et al.* Nondestructive imaging of dielectric-constant profiles and ferroelectric domains with a scanning-tip microwave near-field microscope. *Science* **276**, 2004–2006 (1997).
- Gao, C. & Xiang, X.-D. Quantitative microwave near-field microscopy of dielectric properties. *Rev. Sci. Instrum.* **69**, 3846–3851 (1998).
- Emery, V. J. & Kivelson, S. A. Frustrated electronic phase separation and high-temperature superconductors. *Physica C* **209**, 597–621 (1993).
- Emery, V. J. & Kivelson, S. A. Collective charge transport in high temperature superconductors. *Physica C* **235–240**, 189–192 (1994).
- Emery, V. J. & Kivelson, S. A. Charge ordering in high-temperature superconductors. *Physica C* **263**, 44–48 (1996).
- Salkola, M. L., Emery, V. J. & Kivelson, S. A. Implications of charge ordering for single-particle properties of high- T_c superconductors. *Phys. Rev. Lett.* **77**, 155–158 (1996).
- Zhang, S.-C. A unified theory based on $\text{SO}(5)$ symmetry of superconductivity and antiferromagnetism. *Science* **275**, 1089–1096 (1997).
- Wells, B. O. *et al.* Incommensurate spin fluctuations in high-transition temperature superconductors. *Science* **277**, 1067–1071 (1997).
- Aeppli, G., Mason, T. E., Hayden, S. M., Mook, H. A. & Kulda, J. Nearly singular magnetic fluctuations in the normal state of a high- T_c cuprate superconductor. *Science* **278**, 1432–1435 (1997).
- Mook, H. A. *et al.* Spin fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. *Nature* **395**, 580–582 (1998).
- Moreo, A., Yunoki, S. & Dagotto, E. Phase separation scenario for manganese oxides and related materials. *Science* **283**, 2034–2040 (1999).

22. Kuwahara, H. *et al.* Striction-coupled magnetoresistance in perovskite-type manganese oxides. *Science* **272**, 80–82 (1996).
23. Chen, C. H. & Cheong, S.-W. Commensurate to incommensurate charge ordering and its real-space images in $\text{La}_{0.5}\text{Ca}_{0.5}\text{MnO}_3$. *Phys. Rev. Lett.* **76**, 4042–4045 (1996).
24. Chen, C. H., Cheong, S.-W. & Hwang, H. Charge-ordered stripes in $\text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ with $x > 0.5$. *J. Appl. Phys.* **81**, 4326–4330 (1997).
25. Mori, S., Chen, C. H. & Cheong, S.-W. Pairing of charge-ordered stripes in $(\text{La}, \text{Ca})\text{MnO}_3$. *Nature* **392**, 473–476 (1998).
26. Maezono, R., Ishihara, S. & Nagaosa, N. Phase diagram of manganese oxides. *Phys. Rev. B* **58**, 11583–11596 (1998).
27. Tokura, Y. & Nagaosa, N. Orbital physics in transition-metal oxides. *Science* **288**, 462–468 (2000).
28. Yoo, Y.-K. *et al.* Evidence of strong correlation between doping induced high temperature electronic orders and low temperature magnetic orders in the continuous phase diagram of $\text{Nd}_{1-x}\text{Sr}_x\text{MnO}_3$. *Appl. Phys. Lett.* (submitted).
29. Sachdev, S. Quantum criticality: competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
30. Lide, D. R. *CRC Handbook of Chemistry and Physics* 79th edn, Ch. 12 (CRC, Cleveland, Ohio, 1997).

Acknowledgements

This work was supported by Advanced Energy Projects Division, Office of Computational and Technology Research, US Department of Energy and DARPA.

Correspondence and requests for materials should be addressed to X.-D.X. (e-mail: xdxiang@lbl.gov).

Stress transmission through a model system of cohesionless elastic grains

Miguel Da Silva & Jean Rajchenbach

Laboratoire des Milieux Désordonnés et Hétérogènes (UMR 7603 CNRS), Case 86, Université P. et M. Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France

Understanding the mechanical properties of granular materials is important for applications in civil and chemical engineering, geophysical sciences and the food industry¹, as well as for the control or prevention of avalanches and landslides². Unlike continuous media, granular materials lack cohesion, and cannot resist tensile stresses. Current descriptions of the mechanical properties of collections of cohesionless grains have relied either on elasto-plastic models classically used in civil engineering³, or on a recent model involving hyperbolic equations^{4,5}. The former models suggest that collections of elastic grains submitted to a compressive load will behave elastically. Here we present the results of an experiment on a two-dimensional model system—made of discrete square cells submitted to a point load—in which the region in which the stress is confined is photoelastically visualized as a parabola. These results, which can be interpreted within a statistical framework, demonstrate that the collective response of the pile contradicts the standard elastic predictions and supports a diffusive description of stress transmission. We expect that these findings will be applicable to problems in soil mechanics, such as the behaviour of cohesionless soils or sand piles.

The point-punch test probes the mechanical behaviour of the material. Provided that the partial-differential equation describing the stress field throughout the discrete medium is linear, as is the case in all currently competing models, the response to the point load identifies with the Green's function for the stress equation. We prepared the sample by cutting a 10-mm-thick elastomeric plate into identical square grains (15 mm × 15 mm), whose vertices are trimmed. The elastic grains are then packed into contact and bounded by a metal frame. The Young's modulus and Poisson coefficient of the polycarbonate elastomer are respectively $Y = 3.15 \times 10^5$ Pa and $\nu = 0.36$. The punch consists of a steel sphere 3 mm in diameter and with Young's modulus $Y = 2 \times 10^{11}$ Pa. The bidimensional packing is partially disordered: each linear array,

constituting a layer, is regular, but there is a varying shift with the underlying layer (Fig. 1). Owing to tip effects, the real contacts experienced by each square cell are preferentially located at its vertices. Hence, the present tiling is convenient to model the stress propagation through different disordered contact networks, simply by varying the shift between adjacent layers. The sample is then positioned between two polarizers at right angles, in order to perform photoelastic visualization of the strain state.

Figure 1a shows a photoelastic view of the point-loaded packing: the load P is typically 20 N. Figure 1b shows the active contact network from the location of brightened contacts. In Fig. 1a, the arrow labelled 1 points to a typical compressive contact, for which we recognize a set of quasi-circular bright fringes, whereas the arrow labelled 2 points to a typical shear contact, for which bright lobes are inclined at 45° relative to the interfacial plane.

When varying the packing structure we obtain different patterns for the internal stress. It is informative to perform an ensemble average of such stress patterns. Figure 2 presents the average of 10 digitized images corresponding to different packings. It is clear that the strained region is bounded by a parabola. This is a remarkable result, because, for our model system, it refutes both the usual elastic approach and the alternative hyperbolic proposal. Instead, it supports a diffusive description of the stress transmission, accounted for by equations of the parabolic type.

In classical engineering textbooks, static cohesionless piles made up of elastic grains and submitted to a purely compressive external load are considered to be elastic below the plastic threshold^{3,6}. In the elastic theory, the stress field obeys the biharmonic equation^{7,8}, which is elliptic and implies that the internal state of stress depends on the whole set of boundary conditions. On the other hand, the Mohr–Coulomb yield condition associated with the mechanical equilibrium equations leads to a set of hyperbolic partial-differential equations for the stresses in the plastic state⁶. In the latter case, stresses propagate as waves, and they are only sensitive to some particular points located on the boundaries.

If the results carry over to granular systems in general, made up of grains of arbitrary shape and packed in random fashion, this implies that the elastic theory cannot account for the collective response of a compressed collection of elastic grains, in continuum modelling. Indeed, in two dimensions, the elastic response of a semi-infinite medium to a normal point force is as follows⁹. If the origin (of polar

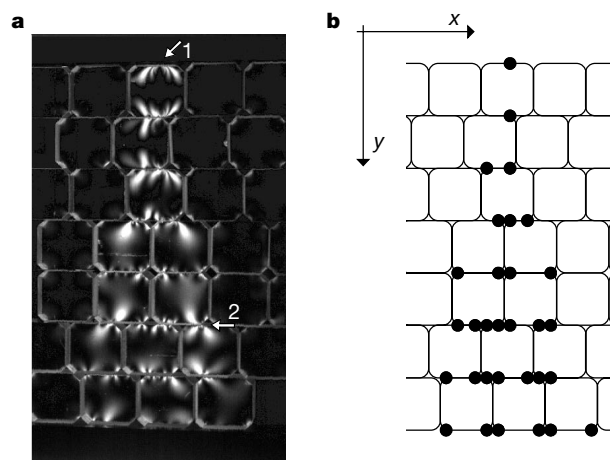


Figure 1 Photoelastic view of a packing submitted to a point load. **a**, Typical photoelastic view of a packing submitted to a point load of 20 N. Each packing is made up of regular arrays of square grains; there is an arbitrary shift between adjacent arrays. Owing to tip effects, the real contacts undergone by the grains are preferentially located at their vertices. The arrow labelled 1 points to a typical compressive contact, and the arrow labelled 2 points to a typical shear contact. **b**, The active contact network drawn on the basis of the location of brightened contacts.

coordinates r, θ) is taken as the point of application of the load P , the stress is everywhere radial, and its magnitude is given by $\sigma_r(r) = (2P/\pi)(\cos \theta/r)$. Hence the curves of constant stress magnitude are a set of circles passing through the point of application of the force, and then a photoelastic visualization displays closed isochromatic (nearly circular) contours also passing through the loading point. In contrast, hyperbolic equations proceeding from plasticity theory⁶, or from the 'BCC model' recently proposed by Bouchaud *et al.*^{4,5}, would imply the existence of two bright rays, arising from the loading point, and propagating along two symmetrically inclined characteristic directions. Therefore our observations oppose all models leading to elliptic or hyperbolic stress equations below the Mohr–Coulomb yield criterion.

Our observations are consistent with the stochastic model of inter-grain force transmission known as the ' q -model', which was devised¹⁰ in order to account for the stress fluctuations in grain piles^{11–15}. In this model, forces are taken to be scalar, and their transmission on the beads beneath is weighted by a random number q uniformly distributed between 0 and 1. Assuming the probability distribution function of the force to be stationary in the medium, the authors could successfully determine this distribution function analytically. The force distribution given by this Ansatz appears to be in very good agreement with both experimental^{11–13} and numerical¹⁵ determinations. In the continuum limit, after an

ensemble averaging, the q -model leads to a simple diffusion behaviour of the externally applied point force through the medium¹⁶. This diffusive behaviour leads straightforwardly to a parabolic boundary for the strained region. We also note that the value of the diffusion coefficient (D is half grain size) derived from the q -model is consistent with the value determined from our experimental estimate, which proceeds from the determination of the bounding parabola parameter ($D = 0.45 \pm 0.05$ grain size).

It may seem surprising that such misunderstanding of parabolic and elliptic behaviour has persisted for half a century in mechanical textbooks. We think that it may be partly due to the confusion between 'reversible' and 'elastic' behaviours. Here individual grains behave elastically, and their collective response is reversible, but not elastic. It is nevertheless worth noting that in the case of granular layers submitted to a point load P , the elasticity theory leads to a pressure profile of equation $\sigma_{yy}(x, y) = (2P/\pi)(y^3/(x^2 + y^2)^2)$ in two dimensions ($\sigma_{yy}(x, y) = (3P/2\pi)y^3/(x^2 + y^2)^{5/2}$ in three dimensions) and that the pressure profile is therefore also bell-shaped, like the gaussian profile predicted by the diffusive model. This kinship is certainly responsible for the successes of computations based on elastic finite-element meshing and used in civil engineering.

Our results do not fit elasticity theory and hyperbolic models; instead, they agree with a diffusive pattern. We note that: (1) the inadequacy of the standard elastic description arises from the alteration in the stress transmission introduced by the unilateral character of the contact force, that is, the grain non-cohesiveness. (2) In materials like ore heaps, soils or sand piles, as well as in our model system, forces are transmitted through point contacts, unlike the finite-element modelling of continuous media. (3) An important feature is the existence of a tree-like hierarchical structure for the force transmission downward from the point load. Besides, it is worth recalling that the experimentally supported parabolic behaviour refers to an ensemble average over numerous pile configurations. Both results uphold stochastic descriptions such as the ' q -model'¹⁰. The success of such a simplified model may seem surprising, as it is scalar and therefore cannot ensure the balance of horizontal internal forces. We argue that this last requirement is obeyed when vectorial forces and disordered packings of grains are considered. For the sake of simplicity, we consider an isostatic, frictionless, disordered system¹⁷. An unequal transmission factor q of the vertical force component naturally ensues, which can be unambiguously determined from the geometrical constraints, without violating the force balance (see Fig. 3). Hence, amorphous packings are characterized both by the hierarchical structure of the inter-grain contact forces, and by the absence of geometrical correlation in the packing (over a few grain sizes). This absence of long-range correlation implies a random series of coefficients q along the force transmission tree (coarse-grained over the short

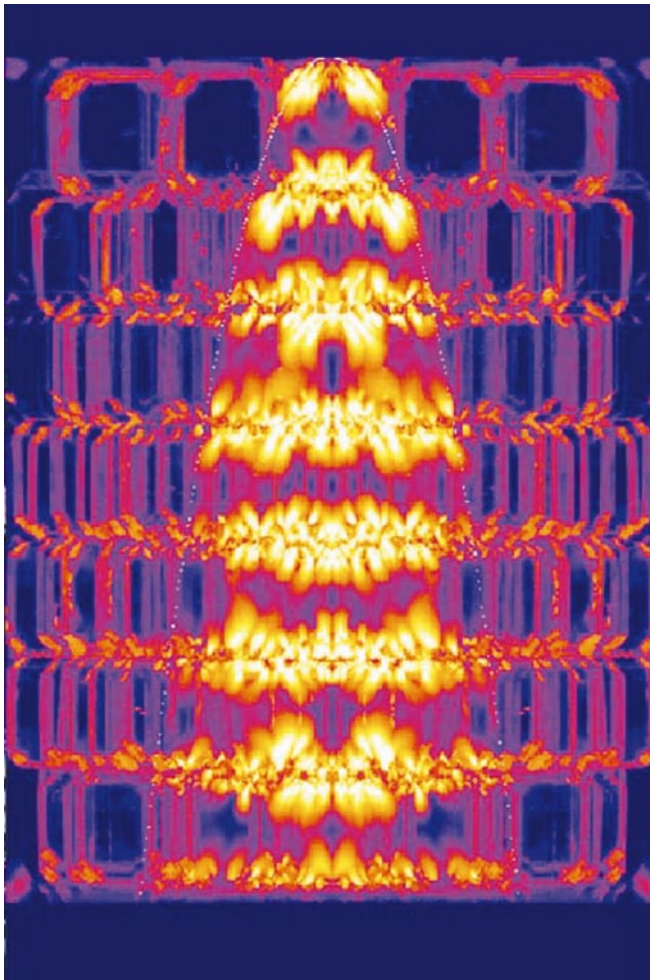


Figure 2 Superimposition of ten digitized pictures, of the type shown in Fig. 1a. In each packing, the shift between horizontal adjacent layers is different. The strained region appears to be bounded by a parabola that we represent by a dashed line, and that is a guide for the eyes. The bounding parabola obeys the equation $x = (2Dy)^{1/2}$, with $D = 0.45 \pm 0.05$ grain size.

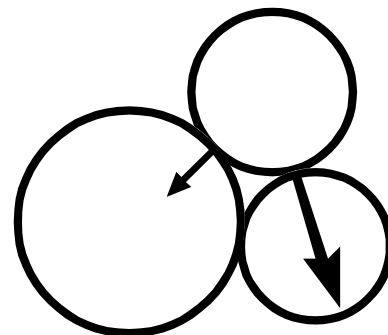


Figure 3 Equilibrium of three frictionless spherical grains. The position disorder naturally yields an unequal transmission factor q on the grains beneath. The q -distribution is shown here to result from the topological disorder of the packing, and is compatible with the force equilibrium.

geometrical correlation length), and consequently a purely diffusive large-scale behaviour for the force. Therefore, we conclude that our results are certainly applicable to disordered packings such as cohesionless soils or sand piles. In contrast, for the particular case of non-random, textured packings, biased and long-range correlated q -series along the force transmission tree are likely to be encountered, thus altering the previous purely diffusive behaviour. □

Received 4 October 1999; accepted 15 June 2000.

- Ennis, B. J. in *Powders and Grains 97* (eds Behringer, R. P. & Jenkins, T. J.) 13–23 (A. A. Balkema, Rotterdam, 1997).
- Gray, J. M. & Hutter, K. Physik granularer Lawinen. *Phys. Bl.* **54**, 37–43 (1998).
- Terzaghi, K. *Theoretical Soil Mechanics* (Wiley, New York, 1943).
- Bouchaud, J. P., Cates, M. E. & Claudin, P. Stress distribution in granular media and nonlinear wave equation. *J. Phys. I (Fr.)* **5**, 639–656 (1995).
- Cates, M. E., Wittmer, J., Bouchaud, J. P. & Claudin, P. Development of stresses in cohesionless poured sand. *Phil. Trans. R. Soc. Lond. A* **356**, 2535–2560 (1998).
- Sokolovskii, V. V. *Statics of Granular Media* (Pergamon, New York, 1965).
- Timoshenko, S. & Goodier, J. N. *Theory of Elasticity* (McGraw Hill, New York, 1951).
- Landau, L. D. & Lifschitz, E. M. *Theory of Elasticity* (Pergamon, New York, 1980).
- Boussinesq, J. *Application des Potentiels à l'Étude de l'Équilibre et des Mouvements des Solides Élastiques* (Gauthier-Villars, Paris, 1885).
- Coppersmith, S. N., Liu, C. H., Majumdar, S., Narayan, O. & Witten, T. A. Model for force fluctuations in bead packs. *Phys. Rev. E* **53**, 4673–4685 (1996).
- Liu, C. H. *et al.* Force fluctuations in bead packs. *Science* **269**, 513–515 (1995).
- Miller, B., O'Hern, C. & Behringer, R. P. Stress fluctuations for continuous sheared granular packing. *Phys. Rev. Lett.* **77**, 3110–3113 (1996).
- Mueth, D. M., Jaeger, H. M. & Nagel, S. R. Force distribution in granular packing. *Phys. Rev. E* **57**, 3164–3169 (1998).
- Ngadi, A. & Rajchenbach, J. Intermittencies in the compression process of a model granular medium. *Phys. Rev. Lett.* **80**, 273–276 (1998).
- Radjai, F., Jean, M., Moreau, J. J. & Roux, S. Force distribution in dense two dimensional granular systems. *Phys. Rev. Lett.* **77**, 274–277 (1996).
- Peralta-Fabi, R., Malaga, C., Reichtman, R. in *Powders and Grains 97* (eds Behringer, R. P. & Jenkins, T. J.) 227–230 (A. A. Balkema, Rotterdam, 1997).
- Moukarzel, C. Isostatic phase transition and instability in stiff granular materials. *Phys. Rev. Lett.* **81**, 1634–1637 (1998).

Acknowledgements

We thank E. Bringuier for a critical reading of the manuscript.

Correspondence and requests for materials should be addressed to J.R. (e-mail: jer@ccr.jussieu.fr).

A colorimetric sensor array for odour visualization

Neal A. Rakow & Kenneth S. Suslick

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, Illinois 61801, USA

Array-based vapour-sensing devices are used to detect and differentiate between chemically diverse analytes. These systems—based on cross-responsive sensor elements—aim to mimic the mammalian olfactory system^{1–3} by producing composite responses unique to each odorant. Previous work has concentrated on a variety of non-specific chemical interactions^{4–11} to detect non-coordinating organic vapours. But the most odiferous, toxic compounds often bind readily to metal ions. Here we report a simple optical chemical sensing method that utilizes the colour change induced in an array of metalloporphyrin dyes upon ligand binding while minimizing the need for extensive signal transduction hardware. The chemoselective response of a library of immobilized vapour-sensing metalloporphyrin dyes permits the visual identification of a wide range of ligating (alcohols, amines, ethers, phosphines, phosphites, thioethers and thiols) and even weakly ligating (arenes, halocarbons and ketones) vapours. Water

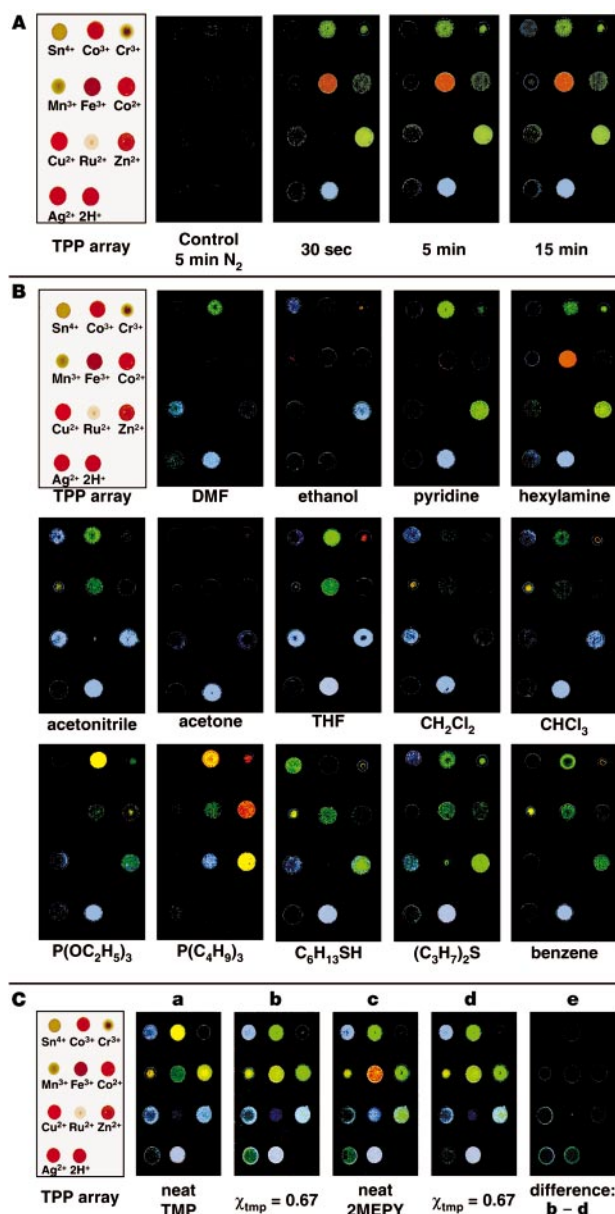


Figure 1 Colour change profiles of a metalloporphyrin sensor array. **A**, Colour change profiles of the metalloporphyrin sensor array as a function of exposure time to n -butylamine vapour. Metalloporphyrins were immobilized on reverse phase silica gel plates. Colour images were obtained with a flatbed scanner (HP Scanjet 3c) at 200 d.p.i. resolution. Subtraction of the initial scan from a scan after 5 min of N_2 exposure was used as a control, giving a black response, as shown. 9.3% n -butylamine in N_2 was then passed over the array and scans made after exposure for 30 s, 5 min and 15 min. The RGB mode images were subtracted (absolute value) using Adobe Photoshop software, with contrast enhancement by expanding the pixel range (a 32-value range was expanded to 256 each for the R, G and B values). **B**, Colour change profiles for a series of vapours; the degree of ligand softness (roughly their polarizability) increases from left to right, top to bottom. Each analyte was delivered to the array as a nitrogen stream saturated with the analyte vapour at 20 °C (to ensure complete saturation, vapour exposures of 15 min or greater were used). DMF, dimethylformamide; THF, tetrahydrofuran. **C**, Two component analysis; saturation responses to mixtures of 2-methylpyridine and trimethylphosphite. Vapour mixtures were obtained by mixing the analyte-saturated N_2 streams at variable flow ratios. A single plate was first exposed to pure trimethylphosphite vapour in N_2 (**a**), followed by increasing mole fractions of 2-methylpyridine up to pure 2-methylpyridine vapour (**c**), followed by decreasing mole fractions of 2-methylpyridine back to pure trimethylphosphite vapour. In both directions, scans were taken at the same mole fraction of trimethylphosphite (χ_{tmp}) and showed excellent reversibility; scans are shown at $\chi_{\text{tmp}} = 0.67$ (**b** and **d**), and their difference map is also shown (**e**).

vapour does not affect the performance of the device, which shows a good linear response to single analytes, and interpretable responses to analyte mixtures. Unique colour fingerprints can be obtained at analyte concentrations below 2 parts per million, and responses to below 100 parts per billion have been observed. We expect that this type of sensing array will be of practical importance for general-purpose vapour dosimeters and analyte-specific detectors (for insecticides, drugs or neurotoxins, for example).

Previous arrays have employed a variety of chemical interaction strategies, including the use of conductive polymers⁴, conductive polymer/carbon-black composites⁵, fluorescent dye/polymer systems^{6,7}, tin oxide sensors^{8,9} and polymer-coated surface acoustic wave devices^{10,11}. Although these systems have demonstrated success in chemical vapour detection and differentiation, their primary aim has been the detection of non-coordinating organic vapours. Many of the most toxic and certainly the most odiferous compounds, however, are excellent ligands for metal ions. Array detection of metal-binding species, such as amines, phosphines and thiols, has been much less explored. (We note that there are two obvious exceptions to the correlation between metal binding and olfactory sensitivity: NO and CO, both of which are produced endogenously as intercellular messenger molecules, and therefore should be olfactorily undetectable¹².) Although very little is known about the structures of the olfactory receptor proteins^{1,2}, we speculate that many of the sensors are likely to contain metal ions at their active site.

Metalloporphyrins are a natural choice for the detection of metal-ligating vapours because of their open coordination sites for axial ligation, their large spectral shifts upon ligand binding, and their intense coloration. Metalloporphyrins have been previously employed for optical detection of gases such as oxygen^{13,14} and ammonia¹⁵, and for vapour detection as chemically interactive layers on quartz crystal microbalances^{16,17}. Here we have taken advantage of the large colour changes induced in metalloporphyrins upon ligand binding, and developed an easy colorimetric technique that minimizes the need for extensive signal transduction hardware. This represents, to our knowledge, the first example of a colorimetric array detector for vapour-phase ligands. Simply by taking the difference before and after exposure of scanned images of the array, we are able to obtain unique colour-change signatures for analytes; these signatures can be used for both qualitative recognition and quantitative analysis.

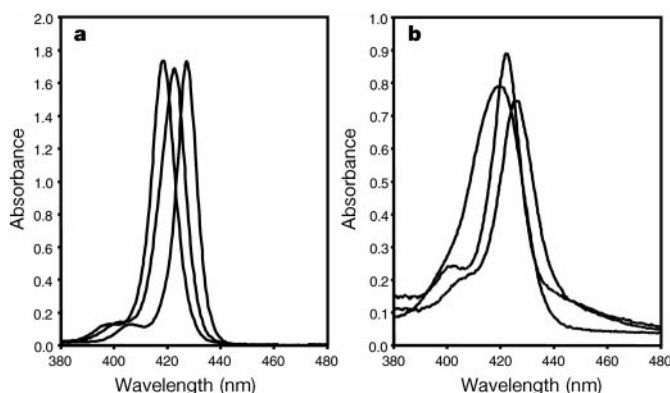


Figure 2 Comparison of Zn(TPP) spectral shifts upon exposure to ethanol and pyridine (py). **a**, In methylene chloride solution; **b**, on the reverse phase support. In both **a** and **b**, the bands correspond, from left to right, to Zn(TPP), Zn(TPP)(C₂H₅OH) and Zn(TPP)(py), respectively. Solution spectra (**a**) were collected using a Hitachi U-3300 spectrophotometer; Zn(TPP), C₂H₅OH and py concentrations were approximately 2 μ M, 170 mM and 200 μ M, respectively. Diffuse reflectance spectra (**b**) were obtained *in situ* with an integrating sphere attachment before exposure to analytes, after exposure to ethanol (5.8% in N₂), and after exposure to pyridine (2.1% in N₂) for 30 min.

Previous studies have elucidated the large spectral changes (and readily observable colour changes) that occur in solution during ligand binding to metalloporphyrins¹⁸. Solution studies have indicated that the magnitude of spectral shift correlates with the polarizability of the ligand¹⁹; hence, there exists an electronic basis for analyte distinction. Using metal centres that span a range of chemical 'hardness' and ligand-binding affinity, a wide range of volatile analytes should be differentiable. Because porphyrins show significant solvatochromic effects, even weakly interacting vapours (for example, arenes, halocarbons or ketones) show distinguishable colorimetric effects.

Solutions of various metalated tetraphenylporphyrins in either methylene chloride or chlorobenzene were spotted onto reverse phase silica thin-layer-chromatography plates to produce the sensor array shown in Fig. 1. The chosen metalloporphyrins have excellent chemical stability on the solid support, and most have well-studied solution ligation chemistry. A stainless-steel flow cell equipped with a quartz window held the plate in a fixed position on an inexpensive flatbed scanner, and permitted exposure of the plate to nitrogen carrying the vapour analyte of interest. Scanning after analyte exposure and subtracting new images from the original scan produces a colour change profile, such as that shown for *n*-butylamine in Fig. 1A. (Details of the image processing are given in Fig. 1 legend.) Virtually all porphyrins are saturated after 30 seconds of exposure, yielding a unique colour fingerprint.

In order to demonstrate the generality of this sensing technique, the vapours were chosen to represent a wide range of chemical functionalities. A comparison of colour changes at saturation for a wide range of analytes is shown in Fig. 1B. Each analyte is easily distinguished from the others, and there are family resemblances among chemically similar species (for example, pyridine and *n*-hexylamine). Analyte distinction originates both in the metal-specific ligation affinities and in their specific, unique colour changes upon ligation.

The metalloporphyrin array can also be used to quantify single analytes and to identify vapour mixtures. Because the images' colour-channel data (that is, RGB values) vary monotonically with porphyrin concentration, we were able to quantify single porphyrin responses to different analytes. Colour-channel data were collected for individual spots and plotted as the quantity $(R_{\text{plt}} - R_{\text{spt}})/R_{\text{plt}}$, where R_{plt} was the red-channel value for the initial silica surface and R_{spt} the average value for the spot. A good linear response is observed until saturation for each porphyrin spot. For example, Zn(TPP), where TPP is 5,10,15,20-tetraphenylporphyrinate(-2), responded linearly to *n*-octylamine between 0 and 0.5 parts per million (p.p.m.). Other porphyrins showed linear response ranges that varied with ligand affinity. Quantifiable, reversible responses are observed even for analyte mixtures of strong ligands, such as pyridines and phosphites (Fig. 1C). Colour change patterns for the mixtures are distinct from either of the neat vapours. Good reversibility was demonstrated for this analyte pair as the vapour mixtures were cycled between the

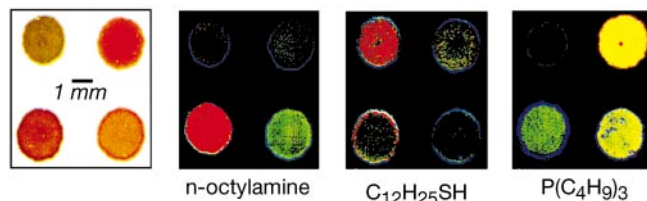


Figure 3 Colour fingerprints at low levels of analyte. The response of a minimized array of four metalloporphyrins (Sn(TPP)(Cl₂), Co(TPP)(Cl), Zn(TPP) and Fe(TPP)(Cl), clockwise from the upper left) is shown for *n*-octylamine, dodecanethiol and tri-*n*-butylphosphine at 1.8 p.p.m.

neat analyte extremes, as shown in Fig. 1C. Response curves of individual porphyrins with varying analyte composition provide a multicomponent analysis of mixtures. For example, for the trimethylphosphite/2-methylpyridine system, Co(TPP) responded linearly in this range to the mole fraction of trimethylphosphite, enabling quantification of an unknown blend of the two vapours.

In an effort to understand the origin of the colour changes upon vapour exposure, diffuse reflectance spectra were obtained for single porphyrin spots before and after exposure to analyte vapours. Porphyrin solutions were spotted in 50- μ l aliquots onto a plate and allowed to dry under vacuum at 50 °C. Spectral shifts on the plate upon analyte exposure correlated well with those seen from solution ligation. For example, exposure of Zn(TPP) to ethanol and pyridine gave shifts very similar to those resulting from ligation in methylene chloride solutions (Fig. 2).

Reverse phase silica was initially chosen as a non-interacting dispersion medium for the metalloporphyrin array, as well as a suitable surface for diffuse reflectance spectral measurements. The hydrophobic nature of the support also greatly reduces interference from water vapour. For instance, a colour fingerprint generated from exposure of the array to *n*-hexylamine vapour (0.86% in N₂) was identical to that for *n*-hexylamine spiked heavily with water vapour (0.43% *n*-hexylamine plus 1.2% H₂O in N₂); see Supplementary Information for further details. The ability to easily detect species in the presence of a large water background represents a substantial advantage over mass-sensing techniques or methodologies that employ polar polymers as part of the sensor array.

Our array's responses come from selective and specific interactions between the analytes and the metalloporphyrin library. This is in marked contrast to the use of a single indicating fluorophore doped in a variety of matrices, where weak matrix-analyte interactions are required to provide selectivity^{6,7}. An advantage of utilizing chemoselective sensors in an array is that unique patterns can be identified at low vapour concentrations. As shown in Fig. 3, an array of four metalloporphyrins can very easily distinguish *n*-octylamine, dodecanethiol and tri-*n*-butylphosphine at 1.8 p.p.m. in N₂. Colour changes of metalloporphyrins on reverse phase silica are slow at analyte exposures below 1 p.p.m. owing primarily to the time necessary to equilibrate the high surface area of the silica surface. As shown in Fig. 4, deposition of the metalloporphyrin library as films on Teflon provides an easy route to miniaturization and gives greatly improved response times. Deposition of spots of approximately 500- μ m on a side present no difficulty; at this size a 20 $\times$ 20 array can be interfaced to an inexpensive CCD (charge-coupled device) detector for device fabrication. Deposition by ink-jet

printing of these films should permit ready, and reproducible, production of chemoselective sensor arrays.

An important goal is the shape-selective distinction of analytes (for example, *n*-hexylamine versus cyclohexylamine). Functionalized metalloporphyrins that limit steric access to the metal ion^{20–22} are candidates for such differentiation. For instance, we have been able^{23,24} to control ligation of various nitrogenous ligands to dendrimer-metalloporphyrins and siloxyl-metalloporphyrins, and induce selectivities over a range of up to 10⁷. The initial results from a small library of siloxyl-metalloporphyrins reveal striking differences between analytes as similar as *n*-octylamine, cyclohexylamine and dipropylamine. Such shape-selective analyte discrimination should add a second tier of molecular recognition to the metal-based array.

The technique that we report here can differentiate between a wide range of species. Such 'smell-seeing' could have a number of different applications, including disposable general-purpose vapour dosimeters as well as analyte-specific detectors. The ability to target metal-binding species provides a complementary technique to currently developing methodologies for non-ligating organic vapours. We are at present working to optimize the array components, complete device fabrication, extend such analysis to liquids, and automate the pattern recognition. □

Methods

Porphyrin preparation

5,10,15,20-Tetraphenylporphyrin, H₂TPP, and its metal complexes were prepared using published methods^{25–28}, and fully characterized by fast-atom-bombardment mass spectrometry, elemental analysis, and ultraviolet–visible and NMR spectroscopies. 5,10,15,20-Tetrakis(pentafluorophenyl)-porphyrinatoiron(III) chloride, Fe(TFP)(Cl), (purchased from Aldrich and used as-received) was added to the porphyrin array because it shows higher ligand-binding constants.

Porphyrin immobilization

The porphyrin array was spotted onto C2 reverse phase silica gel thin-layer-chromatography plates (Whatman KC2, 200 μ m thickness, 350 m² g^{–1} surface area). Porphyrins were spotted from solutions (of a few mM) in chlorobenzene or methylene chloride using a 1- μ l microcapillary tube or microlitre syringe. After spotting, sensor plates were dried under vacuum at 50 °C for 1 h before use.

Vapour exposure

Gas streams containing the vapour of interest were generated by flowing nitrogen at 0.5 l min^{–1} through the neat liquid analyte in a water-jacketed, glass fritted bubbler. All liquid analytes were obtained from Aldrich or Fisher Scientific, and used as-received. Vapour pressures were controlled by regulating the bubbler temperature²⁹. Low p.p.m. levels of *n*-octylamine and tri-*n*-butylphosphine were generated from temperature-regulated analyte/dodecane solutions with the assumption of solution ideality.

Received 25 November 1999; accepted 25 May 2000.

1. Dryer, L. & Berghard, A. Odorant receptors: a plethora of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **20**, 413–417 (1999).
2. Lancet, D. & Ben-Arie, N. Olfactory receptors. *Curr. Biol.* **3**, 668–674 (1993).
3. Axel, R. The molecular logic of smell. *Sci. Am.* **273**, 154–159 (1995).
4. Freund, M. S. & Lewis, N. S. A chemically diverse conducting polymer-based "electronic nose". *Proc. Natl Acad. Sci. USA* **92**, 2652–2656 (1995).
5. Loneragan, M. C. *et al.* Array-based vapor sensing using chemically sensitive, carbon black-polymer resistors. *Chem. Mater.* **8**, 2298–2312 (1996).
6. Walt, D. R. Fiber optic imaging sensors. *Acc. Chem. Res.* **31**, 267–278 (1998).
7. Dickinson, T. A., White, J., Kauer, J. S. & Walt, D. R. A chemical-detecting system based on a cross-reactive optical sensor array. *Nature* **382**, 697–700 (1996).
8. Heilig, A. *et al.* Gas identification by modulating temperatures of SnO₂-based thick film resistors. *Sens. Actuators B* **43**, 45–51 (1997).
9. Gardner, J. W., Shurmer, H. V. & Tan, T. T. Application of an electronic nose to the discrimination of coffees. *Sens. Actuators B* **6**, 71–75 (1992).
10. Crooks, R. M. & Ricco, A. J. New organic materials suitable for use in chemical sensor arrays. *Acc. Chem. Res.* **31**, 219–227 (1998).
11. Grate, J. W. & Abraham, M. H. Solubility interactions and the design of chemically selective sorbent coatings for chemical sensors and arrays. *Sens. Actuators B* **3**, 85–111 (1991).
12. Gelperin, A., Flores, J. & Raccuia-Behling, F. Nitric oxide and carbon monoxide modulate oscillations of olfactory interneurons in a terrestrial mollusk. *J. Neurophysiol.* **83**, 116–127 (2000).
13. Baron, A. E., Danielson, J. D. S., Gouterman, M., Wan, J. R. & Callis, J. B. Submillisecond response times of oxygen-quenched luminescent coatings. *Rev. Sci. Instrum.* **64**, 3394–3402 (1993).
14. Lee, W. *et al.* Halogenated platinum porphyrins as sensing materials for luminescence-based oxygen sensors. *J. Mater. Chem.* **3**, 1031–1035 (1993).
15. Vaughan, A. A., Baron, M. G. & Narayanaswamy, R. Optical ammonia sensing films based on an immobilized metalloporphyrin. *Anal. Commun.* **33**, 393–396 (1996).

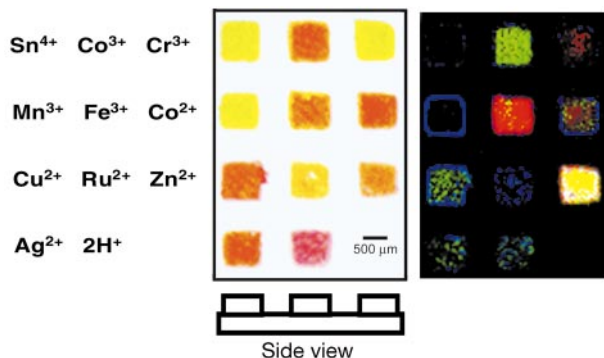


Figure 4 Miniaturized array. The metalloporphyrins were dissolved in a polymer film (dibutylphthalate in polystyrene) and deposited on Teflon; the miniaturization greatly reduces response times. On the left is the array before vapour exposure; on the right is the colour change profile after exposure to *n*-butylamine (1 min, 9.3% in N₂).

16. Brunink, J. A. J. *et al.* The application of metalloporphyrins as coating material for quartz microbalance-based chemical sensors. *Anal. Chim. Acta* **325**, 53–64 (1996).
17. Di Natale, C. *et al.* The exploitation of metalloporphyrins as chemically interactive material in chemical sensors. *Mater. Sci. Eng. C* **5**, 209–215 (1998).
18. Blauer, G. & Sund, H. (eds) *Optical Properties and Structure of Tetrapyrroles* (de Gruyter, Berlin, 1985).
19. Nappa, M. & Valentine, J. S. The influence of axial ligands on metalloporphyrin visible absorption spectra. Complexes of tetraphenylporphinatozinc. *J. Am. Chem. Soc.* **100**, 5075–5080 (1978).
20. Suslick, K. S. & Van Deussen-Jeffries, S. in *Comprehensive Supramolecular Chemistry* (ed. Lehn, J. M.) 141–170 (Elsevier, Oxford, 1996).
21. Bhyrappa, P., Young, J. K., Moore, J. S. & Suslick, K. S. Dendrimer porphyrins: synthesis and catalysis. *J. Am. Chem. Soc.* **118**, 5708–5711 (1996).
22. Chou, J.-H., Nalwa, H. S., Kosal, M. E., Rakow, N. A. & Suslick, K. S. in *The Porphyrin Handbook* (eds Kadish, K., Smith, K. & Guillard, R.) 43–132 (Academic, New York, 2000).
23. Bhyrappa, P., Vijayanthimala, G. & Suslick, K. S. Shape-selective ligation to dendrimer-metallporphyrins. *J. Am. Chem. Soc.* **121**, 262–263 (1999).
24. Sen, A. & Suslick, K. S. Shape selective discrimination of small organic molecules. *J. Am. Chem. Soc.* (in the press).
25. Adler, A. D. *et al.* A simplified synthesis for meso-tetraphenylporphyrin. *J. Org. Chem.* **32**, 476 (1967).
26. Adler, A. D., Longo, F. R., Kampas, F. & Kim, J. On the preparation of metalloporphyrins. *J. Inorg. Nucl. Chem.* **32**, 2443–2445 (1970).
27. Barley, M., Becker, J. Y., Domazetis, G., Dolphin, D. & James, B. R. Synthesis and redox chemistry of octaethylporphyrin complexes of ruthenium(II) and ruthenium(III). *Can. J. Chem.* **61**, 2389–2396 (1983).
28. Datta-Gupta, N. & Bardos, T. J. Synthetic porphyrins II: preparation and spectra of some metal chelates of para-substituted-meso-tetraphenylporphyrins. *J. Pharm. Sci.* **57**, 300–304 (1968).
29. Yaws, C. L. *Handbook of Vapor Pressure* (Gulf, Houston, 1994).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the US NIH and in part by the US DOD and DOE.

Correspondence and requests for materials should be addressed to K.S.S. (e-mail: ksuslick@uiuc.edu).

Timing of the Last Glacial Maximum from observed sea-level minima

Yusuke Yokoyama^{*†}, Kurt Lambeck^{*}, Patrick De Deckker[‡], Paul Johnston^{*} & L. Keith Fifield[§]

^{*} Research School of Earth Sciences, [‡] Department of Geology, [§] Department of Nuclear Physics, Research School of Physical Sciences and Engineering, The Australian National University, Canberra, ACT 0200, Australia

During the Last Glacial Maximum, ice sheets covered large areas in northern latitudes and global temperatures were significantly lower than today. But few direct estimates exist of the volume of the ice sheets, or the timing and rates of change during their advance and retreat^{1,2}. Here we analyse four distinct sediment facies in the shallow, tectonically stable Bonaparte Gulf, Australia—each of which is characteristic of a distinct range in sea level—to estimate the maximum volume of land-based ice during the last glaciation and the timing of the initial melting phase. We use faunal assemblages and preservation status of the sediments to distinguish open marine, shallow marine, marginal marine and brackish conditions, and estimate the timing and the mass of the ice sheets using radiocarbon dating and glacio-hydrostatic modelling. Our results indicate that from at least 22,000 to 19,000 (calendar) years before present, land-based ice volume was at its maximum, exceeding today's grounded ice sheets by $52.5 \times 10^6 \text{ km}^3$. A rapid decrease in ice volume by about 10% within a few hundred years terminated the Last Glacial Maximum at $19,000 \pm 250$ years.

[†] Present address: Space Sciences Laboratory, University of California, Berkeley, California, USA, and Lawrence Livermore National Laboratory, 7000 East Avenue, PO Box 808, L-202 Livermore, California 94550, USA.

The broad and shallow continental margin of northern Australia includes several local bathymetric depressions, the largest of which is the Bonaparte Gulf. During times of sea-level lowstands much of the shelf was exposed, and the sediments deposited in the depressions were protected from wave action by the outer shelf edge. A total of 23 gravity cores, 10 vibrocores and 3 grab samples were collected from present-day water depths of 34 to 147 m along transects across the shelf margin and the Bonaparte depression (Fig. 1a). All cores were sedimentologically examined and a number were selected for micropalaeontological analysis and radiocarbon dating. Depending on faunal assemblages and preservation status (Fig. 1b, see also Methods section below) four distinct bio-sedimentary facies have been recognized: open marine, shallow marine, marginal marine, and brackish water, denoted by OM, SM, MM and BR, respectively, in Fig. 1b. We obtained 41 radiocarbon dates of foraminifera and bivalve molluscs using accelerator mass spectrometry (AMS) techniques and dated some of the larger bivalves using conventional liquid scintillation counting methods. All AMS-dated samples were severely etched (about 40% to 50%) to discard outer shell material that may have been contaminated by secondary carbonate precipitation. Figure 1b summarizes the results for 7 of the cores from water depths between 128 and 95 m. All carbonate ages have been corrected for reservoir effect (400 years; refs 3, 4) and calibrated to a calendar timescale^{5,6}. The cores indicate excellent preservation of faunal specimens, many without evidence for reworking, and the ¹⁴C dates indicate only rare instances of age inversions.

Last Glacial Maximum (LGM) sea-level indicators are preserved in a number of the cores. To reconstruct a sea-level curve based on micropalaeontological evidence, it is necessary to consider the palaeoenvironmental conditions as a sequence of events assuming that little or no break in sedimentation, and no erosion, has occurred. (If a hiatus does occur it results in missing bio-facies and in reworking of the faunas; and this is not observed.) Thus, through identification of a transition from shallow marginal marine to brackish conditions and then back to shallow marginal marine conditions, the timing of the brackish conditions records the interval of lowest sea level. Once this depth is identified in a single core, other cores with depths either side of the identified low sea-level stand are examined to substantiate the reconstruction. This has been done using core GC5 as the master core because it has the best preserved and dated record for brackish water conditions. No truly lacustrine phase has been recognised in any of the cores, indicating that throughout the LGM the Bonaparte depression remained in open contact with the Timor Sea⁷.

The transition from marginal marine to brackish facies occurs in core GC5 at 21,280 calendar years before present (cal. yr BP) at a depth of 340 cm below the sea floor (Fig. 1b). Sediments below this depth contain marine ostracods and planktonic foraminifers, whereas above this boundary dwarf specimens of the benthic foraminifer *Ammonia beccarii* occur along with other shallow to brackish water indicators such as the benthic foraminifer *Elphidium* spp. and the euryhaline ostracod *Cyprideis australiensis*^{8,9}. Marginal marine and brackish-water conditions existed for about 3,000 years, in agreement with results from cores GC4 and GC6 (Fig. 1b) as well as GB1. Sea level at GC5 was, therefore, a few metres below –121 m (Fig. 2a) in this time interval. Deeper cores, both within the depression (GC1, GC2 and GC3) and outside it, do not indicate any brackish facies and place a lower limit of –125 m on the position of relative sea-level at these locations. The microfossil assemblages of cores GC10 and GC11, at –103 and –101 m respectively, are indicative of a marginal beach or coastal lagoon environment between 17,650 and 17,450 cal. yr BP. In core GC7, undamaged littoral-dwelling bivalves occur at –107 m with an age of 17,610 cal. yr BP (Fig. 1b). Figure 2a illustrates the results for all samples free from post-depositional disturbance. They indicate that between 22,000 and 19,000 cal. yr BP a rapid rise of 10–15 m occurred.

$\cdot s^{10,11}$

(1)

where

(2)

is a measure of the ice volume V_i that contributed to the sea-level rise. We call this the ice-equivalent sea-level change. A_w is the ocean area and ρ_i, ρ_w are the average densities of ice and ocean, respectively. $\Delta\zeta_i^*$ is the total glacio-hydro-isostatic contribution to sea-level change. At the Australian sites, away from the former ice sheets, the isostatic response is one of subsidence of the sea floor and coastal zone because of the meltwater loading of the ocean floor. Here, $\Delta\zeta_i^*$ is

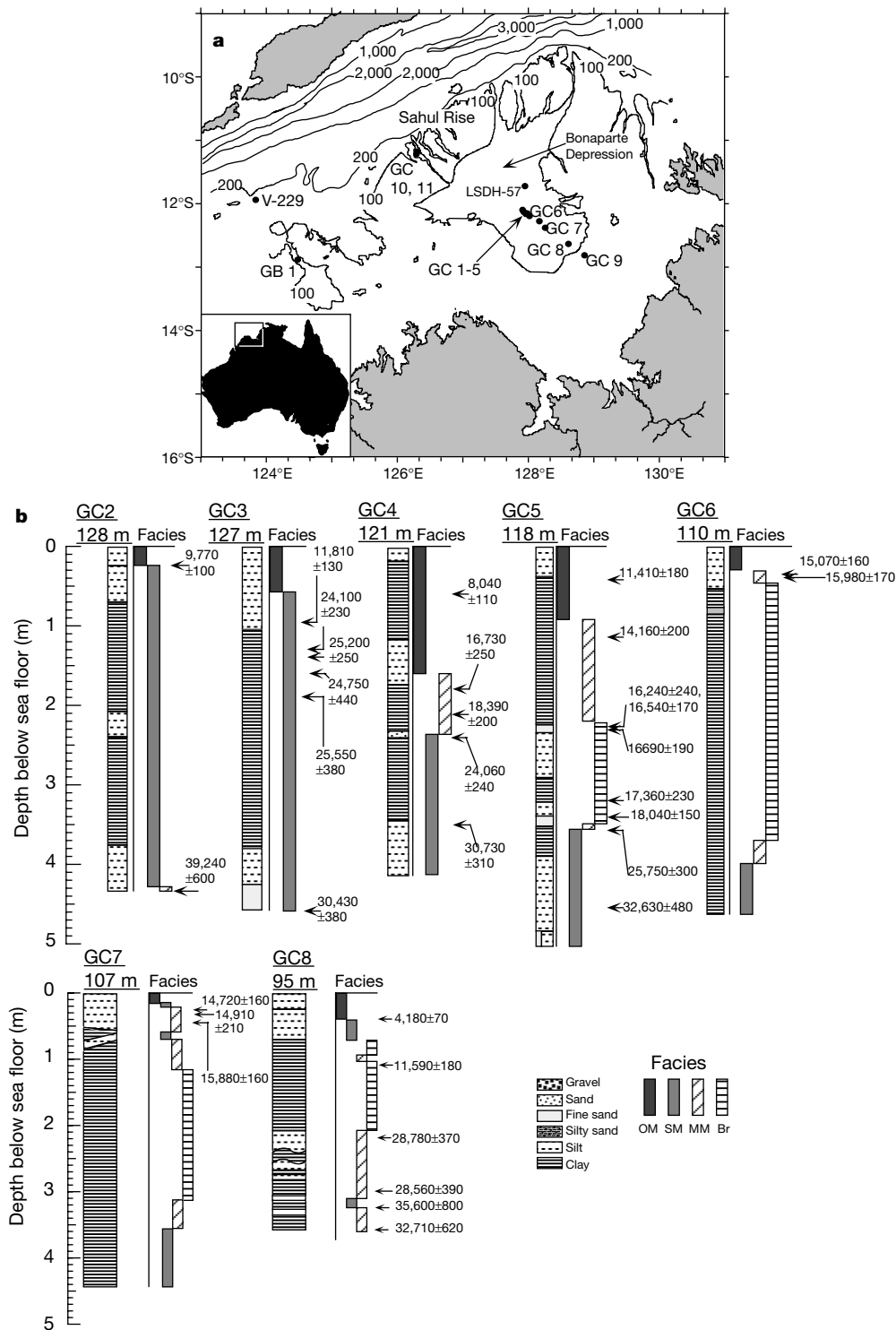


Figure 1 Location and stratigraphy of the Bonaparte cores. **a**, Location of the sampling sites on the northwest shelf of Australia with bathymetric contours (in metres) and coring sites. **b**, Sedimentological description with facies inferred from micro-palaeontological methods (see Methods) and ^{14}C dates for gravity cores. Four facies are

identified as follows: OM, open marine; SM, shallow marine; MM, marginal marine; and BR, brackish-water (see Methods section). Dated horizons are indicated and all ages are in ^{14}C years. The complete data set, including calibrated ages, is available from <http://www.ngdc.noaa.gov/paleo/data.html>.

about 10% of the primary contribution $\Delta\zeta_e$ and the relative sea levels lie above the ice-equivalent sea level at the time of the LGM¹².

The isostatic effects also need to be considered when estimating ice volumes from eustatic sea level $\Delta\zeta_{\text{eust}}$, defined as the globally averaged sea-level change. The value of $\Delta\zeta_i$ in equation (1) averaged over the oceans, $\langle\Delta\zeta_i\rangle_{\text{Aw}}$, at any time is non-zero because of the change in shape and depth of the ocean basin under the changing surface ice and water load. Hence the spatially averaged change in sea level at any time is $\Delta\zeta_{\text{eust}} = \Delta\zeta_e + \langle\Delta\zeta_i\rangle_{\text{Aw}}$ and the last term contributes about 20% to $\Delta\zeta_{\text{eust}}$. In addition to the time dependence of the depth of the ocean basin, the evaluation of the isostatic term in equation (1) and the integral in equation (2) includes the time dependence of both the coastline and the area of shallow sea floor covered by grounded ice.

Figure 2b illustrates the ice-equivalent sea level and associated land-based ice volume, inferred from the Bonaparte Gulf observations. These results use a rigorous glacio-hydro-isostatic formulation with Northern Hemisphere ice models that are consistent with observational evidence from these localities and with an Antarctic ice model that contributes 25 m to $\Delta\zeta_e$ (refs 13–15). Earth model parameters used in these corrections are based on analysis of Australian postglacial sea levels¹² and uncertainties in these parameters are included in evaluating the error bars of Fig. 2b. Because the isostatic corrections are about 10% of the total sea-level change, uncertainties in the distribution of ice between the different centres of glaciation contribute little to the magnitude of the isostatic effects

themselves. The consequence of the isostatic corrections is that the ice-volume-equivalent sea level at the LGM lies between –135 and –130 m, corresponding to a volume of grounded ice in excess of the present volume by $(52 \pm 2) \times 10^6 \text{ km}^3$.

Earlier observations of the LGM sea level on the Australian shelf^{7,16} and elsewhere^{17,18} have shown considerable uncertainty, in part because of the limitation of the preserved indicators and in part because of inadequate age determination. Of the previous data, that of greatest importance is the Barbados evidence where the depth–age relationship of corals have placed lower limits on the local sea-level curve^{19,20}. Figure 2b includes isostatically corrected sea-level estimates—using the same model and model parameters as for the Bonaparte data—for the Barbados data as well as three earlier estimates from northern Australia⁷. The main group of Barbados results, from *Acropora* corals that live in water depths down to about 5 m (ref. 19), occur in the interval from 19,000 to 17,000 cal. yr BP and, when corrected for the differential isostatic effects, are consistent with the Bonaparte Gulf data. The two older data points from Barbados lie at greater depths and are also consistent with the new data for the corresponding period. The agreement between the Barbados and Bonaparte estimates shortly after 19,000 cal. yr BP indicate that the equivalent sea level stood at about –120 m at this time, above the coral and marginal marine samples. The earlier brackish water samples occur at about –135 m equivalent sea level. These estimates are from minimally reworked fauna showing evidence of bleaching and without age inversions. Thus a rapid rise in sea level occurred at about 19,000 cal. yr BP, within a time interval defined by the dating accuracy of a few hundred years. This rapid rise was followed by a period of about 2,000 to 3,000 years of much slower melting of the ice sheets before the onset of the main phase of late-glacial melting. The bulk of the Barbados evidence, therefore, does not correspond to the period of maximum glaciation but to the early stages of the late-glacial phase.

The inferred ice-equivalent sea level of –130 to –135 m is consistent with the glaciological lower limit estimate of –127 m established by the CLIMAP project¹ and does not support the maximum CLIMAP ice-sheet reconstruction of –163 m. (It has been argued that upper- and lower-limit CLIMAP models contain excessive ice when compared with an isostatically corrected Barbados sea-level result of –105 m (ref. 21). But this latter quantity corresponds to $\Delta\zeta_{\text{eust}}$, not $\Delta\zeta_e$. A $\Delta\zeta_{\text{eust}}$ of –105 m corresponds in fact to $\Delta\zeta_i$ of about –135 m.)

The global sea-level observations do not indicate where the ice was stored. Analyses of sea-level data from formerly glaciated regions provide Northern Hemisphere ice volumes that are less than those contained in glaciologically based models^{13,21–23} and the previously noted^{14,24} discrepancy between the Northern Hemisphere and total LGM ice volumes remains. Changes in Antarctic ice may account for part of this, although estimates based on Antarctic rebound analyses are inadequate to explain the entire imbalance²⁵. In the inversion of sea-level data it is the rebound within the former ice margins that is most sensitive to changes in ice volume, but the necessary observations occur only once the area becomes ice free. Hence, ice volumes for the LGM and early part of the late-glacial period remain poorly constrained in such analyses²⁶. The rapid rise in sea level noted at 19,000 cal. yr BP may provide part of the answer to the missing-ice problem: that the LGM ice sheets were initially cold-based, steeply domed and thick but then, in response to basal thawing over some regions, evolved rapidly into relatively thin ice cover with a significant reduction in ice volume^{27,28}. Such a hypothesis is consistent with the geomorphological field evidence²⁷, with glaciological modelling²⁹ and with the ice models inferred from rebound analyses^{13,22}. □

Methods

Facies determination

To define water depth within 1–2 m in the Bonaparte Gulf, it is necessary to examine the sequence of depositional events immediately before and after the deposition of the

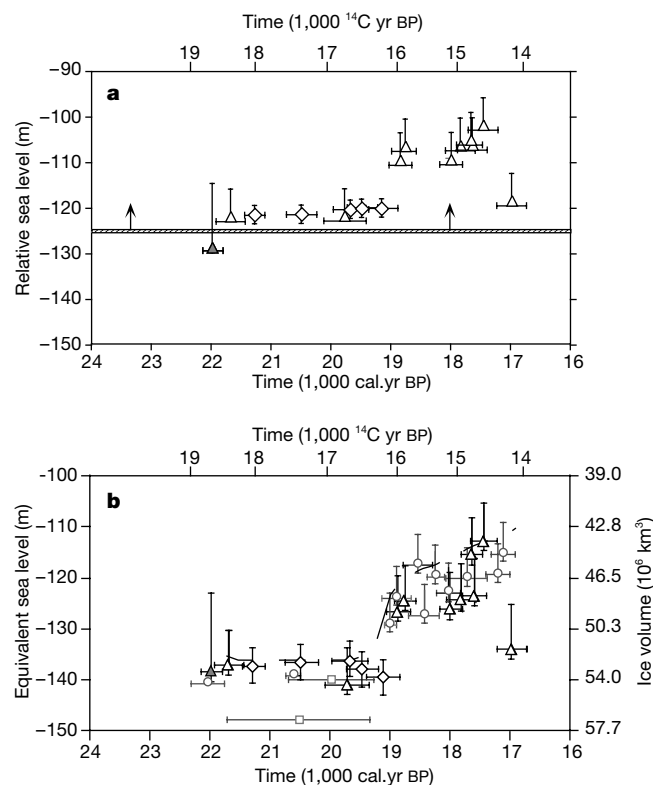


Figure 2 Sea level and ice volumes during the LGM. **a**, Observed sea level for the Bonaparte depression. The diamonds correspond to brackish-water facies, the triangles to marginal marine facies and the filled triangle to shallow marine facies. The hatched zone corresponds to the lower-limit estimate for relative sea level (according to results obtained from GC2 and 3 that neither core contains the shallow-water facies). **b**, Equivalent sea-level and ice-volume estimates. The circles correspond to the isostatically corrected Barbados data^{19,20}, and the squares correspond to earlier estimates for the Northwest Shelf⁷. The diamonds correspond again to brackish-water facies, the triangles to marginal marine facies, and the filled triangle to shallow marine facies. The dashed line defines the ice-equivalent sea-level function. Age uncertainties correspond to 1σ except for U/Th dated Barbados corals (2σ level)²⁰.

brackish facies. In a falling sea-level sequence the following biosedimentological facies sequence can be expected within the core: 1, open marine; 2, shallow marine; 3, marginal marine; and 4, brackish conditions. The last three facies are seen, for example, in the lower part of core 5 (Fig. 1b). When sea level rises the sequence reverts to marginal marine and then to shallow marine. It is this sequence of facies changes before and after the deposition of the brackish-water sediments that determine the position and duration of shallow-water conditions. The four depositional sequences are as follows.

"Open marine" was identified by the presence of fragile pteropod remains, numerous and diversified benthic foraminifer and ostracod taxa, that are indicative of normal marine salinity. In analogy with modern conditions, water depths are around 20 m because planktonic foraminifers are either absent or rare.

"Shallow marine" was characterized by the absence of pteropods and the presence of numerous and well-preserved foraminifer and ostracod taxa (which include species of *Agilaila*, *Argilloecia*, *Callistocythere*, *Loxococoncha*, *Pterygocythereis*, *Uloleberis*, in addition to common bairdiid taxa); there are no planktonic foraminifers. Biodiversity of the microbiota is substantially lower than in the open marine facies. Water depths are around 10 m.

"Marginal marine" was defined by a change to a lower diversity of benthic marine ostracods (commonly encountered species: *Neocytheretta* spp., *Xestoleberis* sp.) and benthic foraminifers, accompanied by the robust endobenthic scaphopods. Most specimens show signs of abrasion. Bivalve molluscs are common and frequently damaged. Broken echinoid spines and bryozoan remains are ubiquitous. Terrigenous material, mostly quartz, is common. Water depths are less than 5 m, being within the zone of tidal influence.

"Brackish water" was identified by a paucity of typically marine organisms and the presence of the foraminifer *Ammonia beccarii*, often in large numbers—and as dwarf morphs due to environmental stresses such as fluctuating salinities and temperatures—accompanied by the euryhaline ostracods *Cyprideis australiensis*, *Leptocythere* spp. and *Neocytheretta* spp. Signs of bleaching of the calcareous shells, indicating exposure to corrosive waters, are common. Terrigenous material is common. Salinities are well below sea-water salinity due to a strong influence of continental waters, such as found in estuaries and tidal flats. Water depths correspond to near mean sea level and up to the high-tide level. Present tidal range is ± 3 m and this represents the upper limit of the accuracy of these sea-level indicators.

Received 23 July 1999; accepted 16 June 2000.

- CLIMAP Project Members. Seasonal reconstructions of the earth's surface at the Last Glacial Maximum. *Geol. Soc. Am. Map Chart Ser. MC-36*, 1–18 (1981).
- Bard, E. Ice age temperatures and geochemistry. *Science* **284**, 1133–1134 (1999).
- Stuiver, M. & Braziunas, T. F. Modeling atmospheric ^{14}C ages of marine samples to 10,000 BC. *Radiocarbon* **35**, 137–189 (1993).
- Edwards, R. L. *et al.* A large drop in atmospheric $^{14}\text{C}/^{12}\text{C}$ and reduced melting in the Younger Dryas, documented with ^{230}Th ages of corals. *Science* **260**, 962–968 (1993).
- Stuiver, M. & Reimer, P. J. Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
- Bard, E., Arnold, M., Hamelin, B., Tisnerat-Laborde, N. & Cabioch, G. Radiocarbon calibration by means of mass spectrometric $^{230}\text{Th}/^{234}\text{U}$ and ^{14}C ages of corals. An updated data base including samples from Barbados, Mururoa, and Tahiti. *Radiocarbon* **40**, 1085–1092 (1998).
- van Andel, T. H., Heath, G. R., Moore, T. C. & McGeary, D. F. R. Late Quaternary history, climate, and oceanography of the Timor sea, Northwestern Australia. *Am. J. Sci.* **265**, 737–758 (1967).
- De Decker, P. An account of the techniques using ostracods in palaeolimnology in Australia. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **62**, 463–475 (1988).
- Yassini, I. & Jones, B. G. (eds) *Recent Foraminifera and Ostracoda from Estuarine and Shelf Environments on the Southeastern Coast of Australia* (Univ. Wollongong Press, 1995).
- Farrall, W. E. & Clark, J. A. On postglacial sealevel. *Geophys. J.* **46**, 79–116 (1976).
- Nakada, M. & Lambeck, K. Glacial rebound and relative sea-level variations: a new appraisal. *Geophys. J. R. Astron. Soc.* **90**, 171–224 (1987).
- Lambeck, K. & Nakada, M. Late Pleistocene and Holocene sea-level change along the Australian coast. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **89**, 143–176 (1990).
- Lambeck, K., Smith, C. & Johnston, P. Sea-level change, glacial rebound and mantle viscosity for northern Europe. *Geophys. J. Int.* **134**, 102–144 (1998).
- Nakada, M. & Lambeck, K. The melting history of the Late Pleistocene Antarctic ice sheet. *Nature* **33**, 36–40 (1988).
- Fleming, K. *et al.* Refining the eustatic sea-level curve since the Last Glacial Maximum using far- and intermediate-field sites. *Earth Planet. Sci. Lett.* **163**, 327–342 (1998).
- Ferland, M. A., Roy, P. S. & Murray-Wallace, C. V. Glacial lowstand deposits on the outer continental shelf of southeastern Australia. *Quat. Res.* **44**, 294–299 (1995).
- Ota, Y., Matsushima, Y. & Moriaki, H. Notes on the Holocene sea-level study in Japan. *Quat. Res. Jpn* **21**, 133–143 (1982).
- Colonna, M., Casanova, J., Dullo, W.-C. & Camoin, G. Sea-level changes and $\delta^{18}\text{O}$ record for the past 34,000 yr from Mayotte Reef, Indian Ocean. *Quat. Res.* **46**, 335–339 (1996).
- Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Bard, E., Hamelin, B. & Fairbanks, R. G. UpTh ages obtained by mass spectrometry in corals from Barbados: sea level during the past 13,000 years. *Nature* **346**, 456–458 (1990).
- Peltier, W. R. Ice age paleotopography. *Science* **265**, 195–201 (1994).
- Tushingham, A. M. & Peltier, W. R. Ice-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of post-glacial relative sea level change. *J. Geophys. Res.* **96**, 4497–4523 (1991).
- Lambeck, K. Limits on the areal extent of the Barents Sea ice sheet in Late Weichselian time. *Glob. Planet. Change* **12**, 41–51 (1996).
- Andrews, J. T. A case of missing water. *Nature* **358**, 281 (1992).
- Zwartz, D., Lambeck, K., Bird, M. & Stone, J. in *The Antarctic Region: Geological Evolution and Processes* (ed. Ricci, C. A.) 821–828 (Terra Antarctica, Siena, 1997).

- Johnston, P. & Lambeck, K. Automatic inference of ice models from postglacial sea-level observations: Theory and application to the British Isles. *J. Geophys. Res.* **13179**–13194 (2000).
- Kleman, J. & Hättestrand, C. Frozen-bed Fennoscandian and Laurentide ice sheets during the Last Glacial Maximum. *Nature* **402**, 63–66 (1999).
- Clark, P. U., Alley, R. B. & Pollard, D. Northern hemisphere ice-sheet influences on global climate change. *Science* **286**, 1104–1111 (1999).
- Licciardi, J. M., Clark, P. U., Jenson, J. W. & Macayeal, D. R. Deglaciation of a soft-bedded Laurentide ice sheet. *Quat. Sci. Rev.* **17**, 427–448 (1998).

Acknowledgements

We thank J. Marshall for providing access to cores collected by the Australian Geological Survey Organisation.

Correspondence and requests for materials should be addressed to K.L. (e-mail: Kurt.Lambeck@anu.edu.au).

Cursoriality in bipedal archosaurs

Terry D. Jones[†], James O. Farlow[‡], John A. Ruben[†], Donald M. Henderson[§] & Willem J. Hillenius^{||}

^{*} Zoology Department, Oregon State University, Corvallis, Oregon 97331, USA

[‡] Department of Geosciences, Indiana-Purdue University, Fort Wayne, Indiana 46805, USA

[§] Department of Cell Biology and Anatomy, School of Medicine, The Johns Hopkins University, Baltimore, Maryland 21205, USA

^{||} Biology Department, College of Charleston, Charleston, South Carolina 29424, USA

Modern birds have markedly foreshortened tails and their body mass is centred anteriorly, near the wings^{1–5}. To provide stability during powered flight, the avian centre of mass is far from the pelvis, which poses potential balance problems for cursorial birds. To compensate, avians adapted to running maintain the femur subhorizontally, with its distal end situated anteriorly, close to the animal's centre of mass; stride generation stems largely from parasagittal rotation of the lower leg about the knee joint^{6–12}. In contrast, bipedal dinosaurs had a centre of mass near the hip joint and rotated the entire hindlimb during stride generation^{4–8,11–13}. Here we show that these contrasting styles of cursoriality are tightly linked to longer relative total hindlimb length in cursorial birds than in bipedal dinosaurs. Surprisingly, *Caudipteryx*, described as a theropod dinosaur^{14,15}, possessed an anterior centre of mass and hindlimb proportions resembling those of cursorial birds. Accordingly, *Caudipteryx* probably used a running mechanism more similar to that of modern cursorial birds than to that of all other bipedal dinosaurs. These observations provide valuable clues about cursoriality in *Caudipteryx*, but may also have implications for interpreting the locomotory status of its ancestors.

In contrast to bipedal dinosaurs, the femur in cursorial birds contributes little to generation of stride length, and avian hindlimb movement is largely the result of retraction of the lower leg^{7–12}. Hence, it might be expected that relative stride length in cursorial birds would be lower than that in bipedal dinosaurs. However, this is probably not the case—lengths of 'effective hindlimb' segments in birds (tibiotarsus + tarsometatarsus) and dinosaurs (femur + tibia + metatarsal III) are equivalent¹⁶ (Fig. 1a). Consequently, total hindlimb length in cursorial birds is invariably one-and-a-half times longer than in theropod and ornithomimid dinosaurs (Fig. 1b). Additionally, as adept avian runners have evolved repeatedly from flighted ancestors¹⁷, we conclude that these profound anatomical modifications that facilitate avian cursoriality are multiple convergent responses to secondary resumption of cursoriality in distantly related taxa independently derived from flighted ancestors.

[†] Present address: Department of Biology, Stephen F. Austin State University, Nacogdoches, Texas 75962, USA.

Caudipteryx has been described as a feathered dinosaur^{14,15} and therefore would be expected to have had a dinosaurian mechanism of cursoriality. However, relative total hindlimb proportions in *Caudipteryx* contrast sharply with those in all other bipedal dinosaurs and are indistinguishable from those in cursorial birds (Fig. 1b). Accordingly, based on the tight linkage of hindlimb proportions to cursorial mechanisms in bipedal archosaurs, we suggest that *Caudipteryx* ran using a mechanism more similar to that of modern cursorial birds than to typical dinosaurs. Significantly, lower leg (tibia + metatarsal) length in *Caudipteryx* is also the same as the 'effective hindlimb' length of cursorial birds, which is equivalent to total hindlimb length in theropods (Fig. 1a).

Because the relative hindlimb proportions of *Caudipteryx* are indistinguishable from those of cursorial birds, it was also likely to have had its centre of mass situated anteriorly, rather than

posteriorly as in other dinosaurs. Importantly, an anterior, cursorial birdlike centre of mass in *Caudipteryx* is demonstrated by application of Henderson's mathematical–computational model for centre of mass in tetrapods¹⁸. This analysis indicates that the centre of mass in *Caudipteryx* was around 2.3 times more anterior to the acetabulum than in the dromaeosaurid theropod *Deinonychus* (Fig. 2).

An anterior centre of mass is consistent with the brevity of the tail of *Caudipteryx*, which is among the shortest, if not the shortest, of all bipedal dinosaurs. Additionally, as in cursorial birds but not dinosaurs, the diminutive tail and the complete absence of a femoral fourth trochanter indicate that the caudofemoralis was not an important locomotory muscle. The likely absence of well developed caudofemoralis musculature indicates that the short tail may have been relatively narrow, with most of its musculature devoted to generating tail movement.

Of all the bipedal dinosaurs, only *Caudipteryx* possessed limb proportions and a centre of mass like those of cursorial birds and, by extension, a cursorial-bird-like running style. These observations might provide valuable clues about the lifestyle of *Caudipteryx* but they may also have implications for interpretation of its taxonomic affinities. We suggest that the anatomical uniqueness of *Caudipteryx* must be consistent with one of the following: (1) *Caudipteryx* was simply an unusual theropod dinosaur whose cursorial ancestors abandoned dinosaurian locomotion and assumed the unique morphology and running style of cursorial birds; (2) *Caudipteryx* was a theropod dinosaur derived from flighted ancestors; (3) *Caudipteryx* was a post-*Archaeopteryx*, secondarily flightless bird and not a 'feathered dinosaur'. The first alternative is supported by cladistic analyses that indicate that *Caudipteryx* was a coelurosauroid theropod^{14,15}. However, none of these analyses has considered *Caudipteryx*'s extensive suite of cursorial-bird-like locomotory characters, and all other theropods were typically dinosaurian in this regard (Figs 1, 2). In this context, it is important to reiterate that the same locomotory specializations present in *Caudipteryx* and cursorial members of eight avian orders are tightly linked to the latter taxa having been derived from flighted ancestors. Accordingly, it is difficult to construct a hypothesis in which the terrestrial theropod ancestors of *Caudipteryx* might have switched to a specialized, running style with an anterior centre of mass, resembling that of

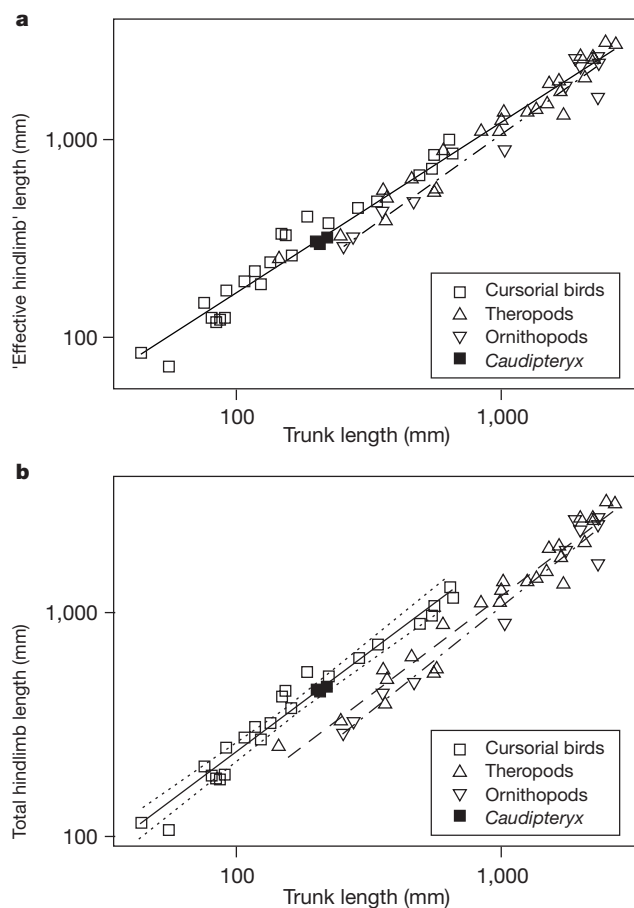


Figure 1 Hindlimb proportions in cursorial birds, theropod and ornithopod dinosaurs, and three specimens of *Caudipteryx*. **a**, The relation of trunk length (T) to 'effective hindlimb' length in bipedal dinosaurs (femur + tibia + metatarsal III), cursorial birds (tibiotarsus + tarsometatarsus) and *Caudipteryx* (tibia + metatarsal III). Cursorial avian and theropod 'effective hindlimb' length equals $0.50T^{0.87}$; for ornithopod dinosaurs it is $0.17T^{0.96}$ ($R^2 = 0.97$, $P < 0.0000$). Extra sum of squares F -test determined that regressions for cursorial bird effective hindlimb length and theropod dinosaur total hindlimb length are indistinguishable ($F_{(2,60)} = 3.654$; $P < 0.05$). Solid line, regression for cursorial birds and theropod dinosaurs; dashed-dotted line, regression for ornithopod dinosaurs. **b**, The relation of trunk length (T) to total hindlimb length in bipedal archosaurs (femur + tibia + metatarsal III in ornithopod and theropod dinosaurs; femur + tibiotarsus + tarsometatarsus in cursorial birds). Hindlimb length equals $0.38T^{0.90}$ for theropod dinosaurs, $0.17T^{0.96}$ for ornithopod dinosaurs, $0.58T^{0.90}$ for cursorial birds ($R^2 = 0.97$, $P < 0.0000$). Solid line, linear regression for cursorial birds. Dotted lines, 99% confidence intervals for cursorial birds. Dashed line, regression for theropod dinosaurs. Dashed-dotted line, regression for ornithopod dinosaurs. Values for *Caudipteryx* specimens were not included in regression calculations. See Methods for details.

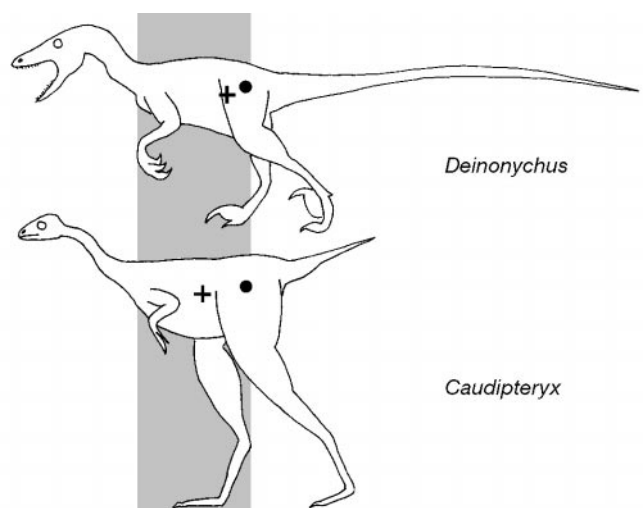


Figure 2 Estimated centre of mass for *Deinonychus* and *Caudipteryx*. Figures are scaled so that trunk lengths are equivalent. Relative to the acetabulum (dot), centre of mass (cross) in *Caudipteryx* approximates centre of mass in cursorial birds and is 2.3 times more anterior than in *Deinonychus*. *Deinonychus* modified from ref. 19. *Caudipteryx* was reconstructed from NGMC 97–4–A and NGMC 97–9–A. For illustrative purposes the femora of *Caudipteryx* have been drawn in a more dinosaurian position. The femora of *Caudipteryx*, as in cursorial birds, was probably more horizontal. See Methods and ref. 18 for details.

cursorial birds, when they were already adept cursors with a posterior centre of mass.

The second proposal, that some theropods were derived from currently unknown, flighted ancestors, has been suggested¹⁹. As noted above, *Caudipteryx*'s locomotory adaptations are consistent with it having been derived from flighted ancestors. However, there is no substantive evidence to support flighted ancestors for any other theropods. Nevertheless, this possibility cannot be dismissed. In the light of problems associated with theories (1) and (2), perhaps the third, that *Caudipteryx* was a secondarily flightless, post-*Archaeopteryx*, cursorial bird, deserves closer scrutiny than it has received so far. We find it a striking coincidence that the only unambiguously feathered theropod was also the only known theropod likely to have utilized locomotory mechanisms identical to those of cursorial birds. □

Methods

Hindlimb and trunk length data

We recorded data on hindlimb element (femur, tibia or tibiotarsus, metatarsal III or tarsometatarsus) and trunk lengths in mature, extant and extinct cursorial (or at least ground-living) birds (from eight orders and 24 genera, including tinamou, cassowary, ostrich, galliforms, roadrunners, bustards, moa and elephantbird) and 38 genera of theropod and ornithomimid dinosaurs (see Supplementary Information). Data were collected from avian and dinosaur specimens in museum collections or obtained from published data and/or scale reconstructions. Individuals were included only if the hindlimb skeleton was adequately known and the trunk was sufficiently known that the describers were able confidently to reconstruct the specimen.

The morphometric data collected from each individual included maximum femoral length, maximum tibia (or tibiotarsus) length (excluding the cnemial crest) maximum metatarsus (or tarsometatarsus) length and trunk length. We defined trunk length as the distance from the first dorsal vertebrae and/or head of the first dorsal rib to the posterior rim of the acetabulum.

Juvenile birds have, for a given trunk length, longer hindlimbs than their adult counterparts²⁰; the same has been hypothesized for tyrannosaurids²¹ and allosaurids²². Similarly, *Bambiraptor feinbergi*, a juvenile dromaeosaurid theropod, has a hindlimb/trunk length ratio of 2.0 (ref. 23) that is comparable to that seen in cursorial birds, but we exclude it from our analyses because of its obviously early stage of development. Additionally, *Sinornithoides youngi* was excluded from this study. The specimen also exhibits a bird-like hindlimb/trunk length ratio, is very small and possessed a cartilaginous sternum²⁴. These observations indicate that this specimen, originally described as 'nearing maturity', may have been more immature than its describers supposed. To avoid confounding ontogenetic variables, the largest individual for each genus (for which we had data) was used in the analysis and data from known immature extant individuals were omitted as were extinct specimens whose maturity has in doubt.

The developmental maturity of *Caudipteryx* (NGMC 97-4-A and NGMC 97-9-A) is indicated by the well ossified sterna, sternal ribs, wrist bones and ankle bones¹⁴. Similar ossification is present in two more recently discovered, equivalently sized specimens, *Caudipteryx dongi* (IVPP V 12344)²⁵ and *Caudipteryx* (uncatalogued IVPP specimen) (Z. Zhou, personal communication), but the former specimen also possesses ossified uncinat processes. As these skeletal elements ossify late in development, there is little doubt that these individuals were mature.

Centre of mass calculations

Lateral and dorsal profiles of *Deinonychus* were compiled from reconstructions in ref. 19. Lateral and dorsal profiles of *Caudipteryx* were reconstructed from NGMC 97-4-A and NGMC 97-9-A. The axial body profiles were mathematically combined to create a solid three-dimensional model from which body volume and the location of centre of mass were calculated (see ref. 18 for details). The lung was conservatively assumed to have been spherical with a volume equal to 10% of body volume and positioned at the anterior portion of the trunk. The density of the lung (0.4 kg m⁻³) was estimated using the ratio of mass-specific lung volume to mass-specific lung parenchyma volume in a large Nile crocodile²⁶.

Received 8 February; accepted 24 May 2000.

1. Pennycuik, C. J. In *The Origin of Birds and the Evolution of Flight* (ed. Padian, K.) 83–98 (California Academy of Sciences, San Francisco, 1986).
2. Norberg, U. B. *Vertebrate Flight* 1–291 (Springer, Berlin, 1990).
3. Roberts, T. J. & Humphreys, H. A. Hind limb forces during acceleration in wild turkeys. *Amer. Zool.* **39**, 65A (1999).
4. Alexander, R. M. Mechanics of posture and gait of some large dinosaurs. *Zool. J. Linn. Soc.* **83**, 1–25 (1985).
5. Alexander, R. M. *Dynamics of Dinosaurs and other Extinct Giants* (Columbia Univ. Press, New York, 1989).
6. Gatesy, S. M. Caudofemoral musculature and the evolution of theropod locomotion. *Paleobiology* **16**, 170–186 (1990).
7. Gatesy, S. M. & Biewener, A. A. Bipedal locomotion: effects of speed, size and limb posture in birds and humans. *J. Zool. Lond.* **224**, 129–147 (1991).
8. Gatesy, S. M. Hind limb scaling in birds and theropods: implications for terrestrial locomotion. *J. Morphol.* **209**, 83–86 (1991).

9. Gatesy, S. M. Guinea fowl function. I: Cineradiographic analysis and speed effects. *J. Morphol.* **240**, 115–125 (1999).
10. Gatesy, S. M. Guinea fowl limb function. II: Electromyographic analysis and motor pattern evolution. *J. Morphol.* **240**, 127–142 (1999).
11. Carrano, M. T. Locomotion in non-avian dinosaurs: integrating data from hindlimb kinematics, *in vivo* strains, and bone morphology. *Paleobiology* **24**, 450–469 (1998).
12. Gatesy, S. M., Middleton, K. M., Jenkins, F. A. Jr & Shubin, N. H. Three-dimensional preservation of foot movements in Triassic theropod dinosaurs. *Nature* **399**, 141–144 (1999).
13. Galton, P. M. The posture of hadrosaurian dinosaurs. *J. Paleontol.* **44**, 464–473 (1970).
14. Ji, Q., Currie, P. J., Norell, M. A. & Ji, S. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
15. Sereno, P. C. The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999).
16. Farlow, J. O., Gatesy, S. M., Holtz, T. R., Hutchinson, J. R. & Robinson, J. M. Theropod locomotion. *Am. Zool.* (in the press).
17. Feduccia, A. *The Origin and Evolution of Birds* 2nd edn (Yale Univ. Press, New Haven, 1999).
18. Henderson, D. M. Estimating the masses and centers of mass of extinct animals by 3-D mathematical slicing. *Paleobiology* **25**, 88–106 (1999).
19. Paul, G. S. *Predatory Dinosaurs of the World* (Simon and Schuster, New York, 1985).
20. Manion, B. L. *The Effects of Size and Growth on the Hindlimb Locomotion in the Chicken*. Thesis, Univ. Illinois at Chicago (1984).
21. Russell, D. A. Tyrannosaurs from the Late Cretaceous of Canada. *National Museum of Canada, National Museum of Natural Sciences Publications in Paleontology* **1**, 1–34 (1970).
22. Foster, J. R. & Chure, D. J. Hindlimb proportion allometry in juvenile and adult *Allosaurus* (Dinosauria, Theropoda). *J. Vert. Paleontol.* **19** (suppl. To 3), 45A (1999).
23. Burnham, D. et al. Remarkable new birdlike dinosaur (Theropoda: Maniraptora) from the upper Cretaceous of Montana. *U. Kansas Paleontol. Contrib.* **13**, 1–14 (2000).
24. Russell, D. A. & Dong, Z. A nearly complete skeleton of a new troodontid dinosaur from the Early Cretaceous of the Ordos Basin, Inner Mongolia, People's Republic of China. *Can. J. Earth Sci.* **30**, 2163–2173 (1993).
25. Zhou, Z. & Wang, X. A new species of *Caudipteryx* from the Yixian Formation of Liaoning, northeast China. *Vertebrata Palasiatica* **38**, 111–127 (2000).
26. Perry, S. 1990. Gas exchange strategy in the Nile crocodile: a morphometric study. *J. Comp. Physiol. B* **159**, 761–769 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. Gatesy, J. Gauthier, N. Geist, F. Hertel, T. Holtz, S. Hunter, J. Hutchinson, L. Martin, M. Norell, S. Olson, T. Roberts, C. Schaff, Z. Zhou, the National Geographic Society and the Carnegie Museum for data, access to specimens, helpful discussions and/or reviews of early versions of this manuscript. This work was supported by NSF grants to J.O.F. and J.A.R. and W.J.H.

Correspondence and requests for materials should be addressed to T.D.J. (e-mail: tdjones@sfasu.edu).

Genetic diversity and disease control in rice

Yuyong Zhu*, Hairu Chen*, Jinghua Fan*, Yunyue Wang*, Yan Li*, Jianbing Chen*, Jinxiang Fan†, Shisheng Yang ‡, Lingping Hu§, Hei Leung||, Tom W. Mew||, Paul S. Teng||, Zonghua Wang|| & Christopher C. Mundt||

* The Phytopathology Laboratory of Yunnan Province, Yunnan Agricultural University, Kunming, Yunnan 650201, China

† Honghe Prefecture Plant Protection Station of Yunnan Province, Kaiyuan 661400, China

‡ Jianshui County Plant Protection Station of Yunnan Province, Jianshui 654300, China

§ Shiping County Plant Protection Station of Yunnan Province, Shiping 662200, China

|| Division of Entomology and Plant Pathology, International Rice Research Institute, MCPO Box 3127, 1271 Makati City, The Philippines

¶ Department of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State University, Corvallis, Oregon 97331-2902, USA

Crop heterogeneity is a possible solution to the vulnerability of monocultured crops to disease^{1–3}. Both theory⁴ and observation^{2,3} indicate that genetic heterogeneity provides greater disease suppression when used over large areas, though experimental data are lacking. Here we report a unique cooperation among farmers,

researchers and extension personnel in Yunnan Province, China—genetically diversified rice crops were planted in all the rice fields in five townships in 1998 and ten townships in 1999. Control plots of monocultured crops allowed us to calculate the effect of diversity on the severity of rice blast, the major disease of rice⁵. Disease-susceptible rice varieties planted in mixtures with resistant varieties had 89% greater yield and blast was 94% less severe than when they were grown in monoculture. The experiment was so successful that fungicidal sprays were no longer applied by the end of the two-year programme. Our results support the view that intraspecific crop diversification provides an ecological approach to disease control that can be highly effective over a large area and contribute to the sustainability of crop production.

Many ecological processes are strongly influenced by spatial scale^{6–9}, causing a major dilemma for experimental biologists, as large-scale field experiments are often prohibitively expensive. For example, there have been increasing calls for ecological approaches to counter the negative environmental impacts of modern agricultural systems^{10,11}. One such approach, the use of within-field crop genetic diversity, has been shown to reduce disease severity in experimental plots and has been used commercially in some

cases^{1–4}. However, experimental procedure and the nature of pathogen dispersal can cause substantial underestimation of the impact of increased diversity on disease in small-scale experimental plots^{2–4}. On the other hand, observations at larger spatial scale are few⁴, and do not allow for unambiguous determination of causal relationships between diversity and disease occurrence.

Our experimental system was blast disease in rice (*Oryza sativa*). Rice is the staple crop for about half of the population of the world¹². The fungus that causes blast disease, *Magnaporthe grisea*, spreads through multiple cycles of asexual conidiospore production during the cropping season, causing necrotic spots on leaves and necrosis of panicles. *M. grisea* interacts on a gene-for-gene basis^{13,14} with its host and has a very varied pathogenesis¹⁵. It exists as a mixture of pathogenic races, that is, genetic variants that attack host genotypes with different resistance genes. Thus, host resistance genes often remain effective for only a few years in agricultural production before succumbing to new pathogenic races^{16,17}.

Our experimental site (Yunnan Province, China) favours the development of rice blast epidemics because of its cool, wet climate. Farmers commonly make multiple foliar fungicide applications to control blast. Glutinous or 'sticky' rice varieties are used for confections and other speciality dishes and have higher market value than other rice types, but have lower yields and are highly susceptible to blast. Non-glutinous, hybrid rice varieties are less susceptible to rice blast and are attacked by a different spectrum of *M. grisea* races. Before 1998, 98% of rice fields in the study area were sown with monocultures of the hybrid rice varieties Shanyuo22 and Shanyuo63. The desirable glutinous varieties were planted in small amounts because of their low yields and vulnerability to blast in this environment. We conducted large-scale tests, made possible through the cooperation of thousands of rice farmers, to determine how the occurrence of rice blast is affected by within-field varietal diversification using mixtures of commonly grown glutinous and hybrid rice varieties. Our approach was based on an observed farmer practice of dispersing single rows of glutinous rice between groups of four rows of hybrid rice at a rate sufficient to meet local demand for glutinous rice (Fig. 1).

In the first year of the experiment, four different mixtures of varieties (Fig. 2) were planted in a 812-ha area consisting of all rice fields in five townships of Shiping County, Yunnan Province. Because of the excellent blast control provided by the variety mixtures, only one foliar fungicide spray was applied. Mixtures were compared to monoculture control plots at 15 survey sites. Unlike standard experiment station fields, control plots of monocultures were small relative to the total area of mixtures planted by farmers in the surrounding area, reducing the potential impact of spore dispersal from the more heavily infected monocultures to the mixture plots^{2–4}. The study was expanded to 3,342 ha of rice fields in 1999. This area consisted of all rice fields in 10 townships that spanned Jianshui and Shiping Counties, with five participating townships and 15 survey sites per county. Procedures were the same as in 1998, except that no foliar fungicide applications were made. In addition, some farmers chose to plant mixtures in a ratio of 1 glutinous: 6 hybrid rows, rather than 1:4.

Diversification had a substantial impact on rice blast severity (Fig. 2). In 1998, panicle blast severity on the glutinous varieties averaged 20% in monocultures, but was reduced to 1% when dispersed within the mixed populations (Fig. 2a). Panicle blast severity on the hybrid varieties averaged 1.2% in monoculture and was reduced to varying degrees in mixed plots, though only the larger differences were statistically significant (Fig. 2b). Results from 1999 were very similar to the 1998 season for panicle blast severity on the susceptible varieties (Fig. 2a), showing that the effect of diversification was very robust among mixtures and between seasons and counties. In contrast, effects of crop diversification on blast severity of the hybrid varieties were larger in 1999 than in 1998. Panicle blast severity on these varieties averaged 2.3% in

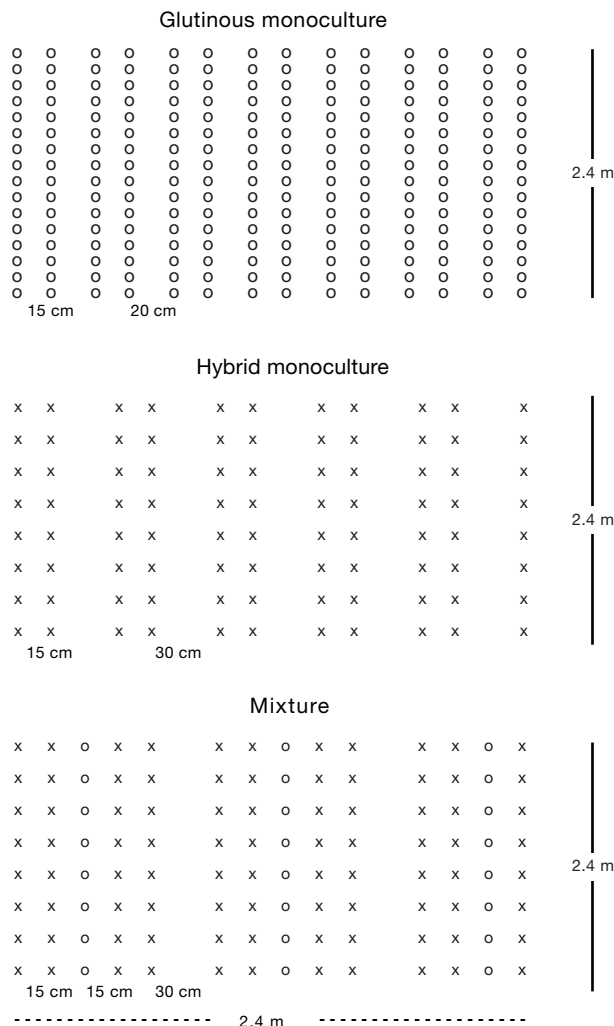


Figure 1 Planting arrangements in rice variety mixture and monoculture survey plots in 1999 and patterned after those used by farmers in Yunnan Province. Each symbol represents a hill of susceptible (O) or resistant (X) rice. Distances between hills within rows were 15 cm for glutinous monocultures, 30 cm for hybrid monocultures and 30 cm for mixtures. Spacings and arrangements were the same in 1998, except that the distance between rows of glutinous rice in monoculture was 13 cm.

monoculture and was reduced to 1.0% in mixed populations (Fig. 2b), despite the fact that hybrids were planted at the same density in both mixture and monoculture survey plots.

Several mechanisms may reduce disease severity in genetically diverse plant populations^{2,4,18}. Increased distance between plant

genotypes, which dilutes inoculum of a given pathogenic race as it is dispersed between compatible host varieties, has been considered the most important mechanism contributing to disease reduction in variety mixtures². Such dilution effects almost certainly had a role in reducing blast disease on the susceptible, glutinous varieties

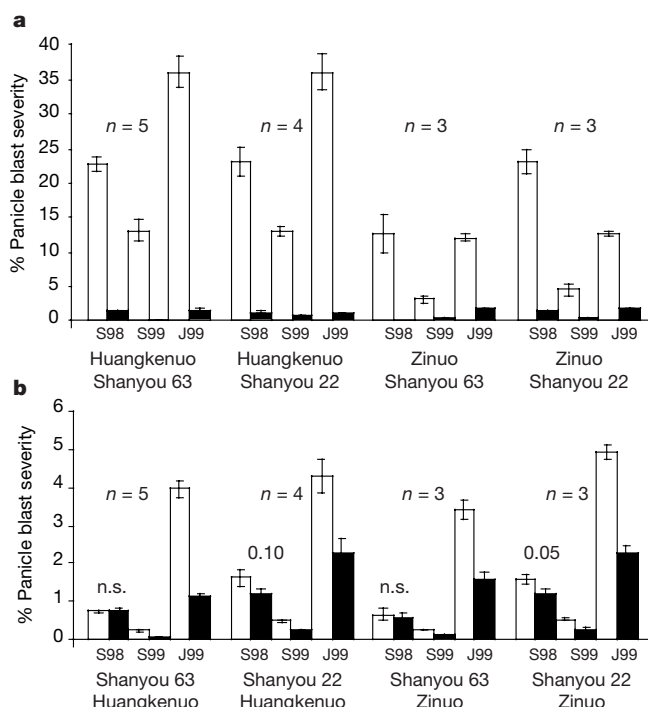


Figure 2 Panicle blast severity (mean percentage of panicle branches that were necrotic due to infection by *Magnaporthe grisea*) of rice varieties planted in monocultures and mixtures. **a**, The susceptible, glutinous varieties Huangkenuo and Zinuo. **b**, The resistant, hybrid varieties Shanyuo22 and Shanyuo63. S98, Shaping County, 1998; S99, Shaping County, 1999; J99, Jianshui County, 1999; open bar, blast severity for a variety grown in monoculture control plots; black bar, blast severity of the same variety when grown in

mixed culture plots in the same fields. Error bars are one s.e.m.; *n*, number of plot means that contribute to individual bars for each of the four combinations of susceptible and resistant variety. All differences between pairs of monoculture and mixture bars are significant at $P < 0.01$ based on a one-tailed *t*-test, unless indicated by 0.05 (significant at $P < 0.05$), 0.10 (significant at $P < 0.10$) or n.s. (not significant at $P = 0.10$).

Table 1 Grain yields and monetary value for rice varieties

Variety or mixture	Hills m ⁻² †	Grain yield ± s.e.m. (Mg per ha)			Crop value (US\$ per ha)		
		Shaping/98	Shaping/99	Jianshui/99	Shaping/98	Shaping/99	Jianshui/99
Huangkenuo	38.1	3.69 ± 0.02	4.07 ± 0.07	5.12 ± 0.05	1291	1424	1794
Shanyuo63	14.8	8.14 ± 0.07	8.41 ± 0.12	9.71 ± 0.07	1709	1765	2039
Mixture	18.5	8.72 ± 0.05	9.53 ± 0.11	10.53 ± 0.12	1912	2166	2341
Huangkenuo	3.7	0.59 (173)	1.19 (300)	0.92 (186)	205	415	323
Shanyuo63	14.8	8.13 (100)	8.34 (99)	9.61 (99)	1707	1751	2018
Huangkenuo	38.1	3.79 ± 0.03	4.15 ± 0.07	5.08 ± 0.10	1328	1452	1778
Shanyuo22	14.8	7.97 ± 0.11	8.12 ± 0.06	9.08 ± 0.20	1673	1705	1907
Mixture	18.5	8.40 ± 0.12	8.77 ± 0.09	10.00 ± 0.16	1838	1941	2231
Huangkenuo	3.7	0.53 (151)	0.71 (177)	0.94 (191)	184	249	330
Shanyuo22	14.8	7.88 (99)	8.06 (99)	9.05 (100)	1654	1692	1901
Zinuo	38.1	3.62 ± 0.04	3.97 ± 0.02	4.90 ± 0.09	1268	1390	1716
Shanyuo63	14.8	8.28 ± 0.13	8.40 ± 0.08	9.63 ± 0.17	1739	1765	2022
Mixture	18.5	8.90 ± 0.22	9.23 ± 0.03	10.46 ± 0.18	1937	2056	2315
Zinuo	3.7	0.48 (146)	0.84 (217)	0.84 (177)	170	294	296
Shanyuo63	14.8	8.42 (102)	8.39 (100)	9.62 (100)	1767	1762	2020
Zinuo	38.1	3.49 ± 0.02	3.82 ± 0.03	4.89 ± 0.11	1220	1337	1711
Shanyuo22	14.8	7.84 ± 0.06	8.14 ± 0.03	9.14 ± 0.05	1646	1710	1919
Mixture	18.5	8.27 ± 0.05	8.86 ± 0.07	9.99 ± 0.03	1807	1965	2227
Zinuo	3.7	0.51 (160)	0.75 (203)	0.92 (193)	178	264	321
Shanyuo22	14.8	7.76 (99)	8.10 (99)	9.08 (99)	1629	1701	1906

The rice varieties were grown as monocultures or mixed in Shaping and Jianshui counties in 1998 and 1999. Crop values based on market prices of 0.21 US\$ per kg for hybrid varieties and 0.35 US\$ per kg for glutinous varieties. Italicized values of hills m⁻², grain yield, and crop value are for the individual varieties within mixtures. Bold values in parentheses are per-hill yields of varieties in mixture expressed as a percentage of per-hill yield of the same variety in monoculture.

* See also Fig. 1.

† In 1998, density of glutinous varieties in monoculture was 40.4 hill m⁻².

in this study. In addition, canopy microclimate data collected at one survey site in 1999 indicate that height differences between the taller glutinous and shorter hybrid varieties resulted in temperature, humidity and light conditions that were less conducive for blast on glutinous varieties in the mixtures than in the monocultures. Disease reductions on hybrid varieties in the mixtures are more difficult to explain. Dilution and microenvironmental modifications are unlikely mechanisms, as the hybrids were planted at the same density in mixtures and monocultures (Fig. 1). The taller glutinous varieties may physically have blocked spore dispersal and/or altered wind patterns compared with the hybrid monocultures. In addition, induced resistance may have some contribution to disease suppression in the hybrids. Induced resistance occurs when inoculation with avirulent pathogen race(s) induces a plant defence response that is effective against pathogen races that would normally be virulent on that host genotype. This has made significant contributions to disease reductions in variety mixtures of other small grain crops^{19,20}.

In 1999, we determined the genetic composition of the pathogen populations derived from inter-planting and monoculture fields using polymerase chain reaction (PCR) fingerprinting²¹ of pathogen isolates. Preliminary results indicate that fields with mixtures supported diverse pathogen populations with no single dominant strain. In contrast, pathogen populations from monoculture fields were dominated by one or a few strains. The more diverse pathogen population from the mixed stands may have contributed to greater induced resistance from incompatible interactions. In the longer term, this increased pathogen diversity may also slow adaptation of the pathogen to resistance genes functioning within a given mixture. Clarifying the mechanisms by which host diversity influenced disease in our study will be helpful in extending these results to other agro-ecosystems. These mechanistic studies are currently underway.

Grain production per hill of glutinous varieties in mixtures averaged 89% greater than that in monoculture (Table 1). As a result, glutinous rice in mixtures produced 18.2% of monoculture yield, on average, though it was planted at rates of only 9.2 and 9.7% that of monoculture in 1998 and 1999, respectively (see also Fig. 1). Reduced disease severity certainly had a role in this yield response, though other factors (for example, improved light interception) may also have had an influence. Despite the increased overall plant density in mixtures (see Fig. 1, bottom), grain yields per hectare of the hybrids in mixture were nearly equal to the corresponding monocultures. Thus, mixed populations produced more total grain per hectare than their corresponding monocultures in all cases (Table 1). Land equivalent ratios²², which estimate the ecological efficiency of mixed populations, indicate that an average of 1.18 ha of monoculture crop land would need to be planted to provide the same amounts of hybrid and glutinous rice as were produced in 1 ha of a mixture (Table 2). After accounting for the differing market values of the two rice types, the gross value per hectare of the mixtures was 14% greater than hybrid monocultures and 40% greater than glutinous monocultures (Table 1).

Though disease reductions are theoretically maximized in random mixtures of plants²³, row mixtures provided the most practical approach in our specific application. As rice is hand-harvested in Yunnan Province, farmers can easily separate the

hybrid and glutinous grains, which are used for different purposes. However, many other approaches can be used to attain within-field genetic diversity of crops^{3,24}. For example, wheat (*Triticum aestivum*) mixtures are grown in the Pacific Northwest of the USA under highly mechanized conditions⁴. In this case, varieties are chosen to be similar in height, maturity and market quality, planted as random mixtures, and harvested and marketed as bulk populations³.

Commercial-scale use of crop diversity has provided observational support for the disease-suppressive effects of crop diversity in a limited number of cases^{4,25,26}, most notably the control of barley (*Hordeum vulgare*) powdery mildew (caused by *Erysiphe graminis*) in the former East Germany²⁶. However, the varietal diversification program in Yunnan Province provided an unusual opportunity to determine causal relationships between crop diversity and disease, as replicated monoculture controls were available for comparison within a substantial expanse of mixed culture. The impact of crop diversification on blast severity in this study was greater than that reported from small-scale experimental plots with this disease³, although we do not have proof that this difference is due only to the spatial scale. By the second year of the project, no foliar fungicides were needed for blast control in the diversified area, though this may not be possible in all seasons. The Yunnan diversification program has resulted in great interest by farmers, and the practice has expanded to more than 40,000 ha in 2000.

The 'Green Revolution' has provided remarkable increases in crop productivity over the past four decades²⁷. However, this agricultural transformation has also resulted in problems, including loss of crop genetic diversity¹¹. The current world population of over six billion does not allow us to return to agricultural production practices of the past. Rather, we need to maintain the benefits of modern agriculture while addressing its drawbacks. In this regard, it is significant that the diversification program described here is being conducted in a cropping system with grain yields approaching 10 Mg ha⁻¹, among the highest in the world. The value of diversity for disease control is well established experimentally and diversity is increasingly being used against wind-dispersed pathogens of small grain cereals⁴. Recent experimental results indicate other applications of diversity, for example, against soil-borne pathogens and for tree crops⁴. The effect of varietal diversification will vary among diseases and agro-ecosystems⁴. Further, one can not expect all variety mixtures to provide functional diversity to a given plant pathogen population^{24,28}, nor can one predict the time for which they may remain effective. Indeed, we have identified variety combinations that provide little or no blast control in Yunnan Province. Nonetheless, our results demonstrate that a simple, ecological approach to disease control can be used effectively at large spatial scale to attain environmentally sound disease control. □

Methods

Study sites

In 1998, townships participating in the diversification experiment were Baxing, Baoxiu, Songchun, Maohe and Xincheng of Shiping County. These townships are contiguous, and all rice fields in the five townships were involved in the diversification program. In 1999, the study area consisted of all rice fields in ten contiguous townships: Chengguang, Dongba, Mianding, Nanzhuang and Xizhuang in Jianshui County; and Baxing, Baoxiu, Songchun, Maohe and Yafanzi in Shiping County.

Disease assessments

To monitor disease, survey plots were established at 15 sites per county, three in each of the five townships participating in the diversification program (15 sites in 1998, 30 sites in 1999). Seedlings were transplanted into the field in April or May in hills of 4–5 plants for glutinous varieties and 1 plant per hill for the hybrid varieties, which produce a greater number of tillers per plant. All plots were managed by farmers and treated in the same manner as the surrounding mixed variety plantings, including fungicide application. In each of the survey sites, a field was divided into three plots. One plot was planted with the mixture grown most commonly by local farmers, and the remaining two plots were monocultures of the glutinous and hybrid variety included in that mixture. For mixtures, the same row spacing of hybrid rice was used as in monoculture, but one row of glutinous rice was added between each group of four rows of hybrid rice, in an 'addition' approach (Fig. 1). Each of the four mixtures was evaluated for disease severity in 3–5 of the 15 survey

Table 2 Land equivalent ratios for rice yield produced in variety mixtures

Mixture	County and year		
	Shiping 1998	Shiping 1999	Jianshui 1999
Huangkenuo/Shanyuo63	1.16	1.28	1.17
Huangkenuo/Shanyuo22	1.13	1.16	1.18
Zinuo/Shanyuo63	1.15	1.21	1.17
Zinuo/Shayuo22	1.14	1.19	1.18

Land equivalent = (yield ha⁻¹ of variety A in mixture/yield ha⁻¹ of variety A in monoculture) + (yield ha⁻¹ of variety B in mixture/yield ha⁻¹ of variety B in monoculture).

sites in each county, depending on the popularity of a given mixture with farmers. Plots ranged from 100 to 450 m² each, depending on field size.

Survey plots were assessed in late August for the severity of blast symptoms, expressed as the percentage of panicle branches that were necrotic due to the effects of *M. grisea*. Disease was assessed at five sampling points in each plot, distributed in a uniform pattern. Twenty hills resulting from the transplanting process were evaluated at each sampling point, with each hill containing about 10 panicles per hill, to give a total of approximately 1,000 panicles evaluated per plot. Each sampled panicle was visually examined by experienced personnel to estimate the percentage of branches that were necrotic due to infection by *M. grisea*. Each panicle was given a rating²⁹ from 0 to 5, where 0 is no disease; 1 is less than 5% of panicle branches necrotic; 2 is 5–30% necrotic; 3 is 30–50% necrotic; 4 is greater than 50% necrotic; and 5 is 100% necrotic. Disease severity was summarized within each plot as $\{[(n_1 \times 1) + (n_2 \times 2) + (n_3 \times 3) + (n_4 \times 4) + (n_5 \times 5)] / \Sigma n_0 \dots n_5\} \times 100$, where $n_0 \dots n_5$ is the number of culms in each of the respective disease categories. Thus, a disease severity of 0% would indicate no disease and 100% would indicate that 100% of panicle branches were necrotic.

Yield evaluation

Plots were hand-harvested, threshed and weighed to determine grain yield. Individual varieties were evaluated separately in mixtures. Land equivalent ratios²² were calculated as (yield ha⁻¹ of variety A in mixture/ yield ha⁻¹ of variety A in monoculture) + (yield ha⁻¹ of variety B in mixture/ yield ha⁻¹ of variety B in monoculture).

Statistical analyses

Each survey plot was considered to be an experimental unit, and analyses were based on mean disease severities and grain yield for each plot. Statistical analyses were conducted separately by year and county owing to differences in disease level. One-tailed *t*-tests were used to determine if blast severity for each of the two varieties in each of the four mixtures differed significantly from its corresponding monoculture control.

Received 18 April; accepted 30 June 2000.

- Browning, J. A. & Frey, K. J. Multiline cultivars as a means of disease control. *Annu. Rev. Phytopathol.* **7**, 355–382 (1969).
- Wolfe, M. S. The current status and prospects of multiline cultivars and variety mixtures for disease resistance. *Annu. Rev. Phytopathol.* **23**, 251–273 (1985).
- Mundt, C. C. in *Rice Blast Disease* (eds Zeigler, R. S., Leong, S. A. & Teng, P. S.) 293–308 (CAB International, Wallingford, 1994).
- Garrett, K. A. & Mundt, C. C. Epidemiology in mixed host populations. *Phytopathology* **89**, 984–990 (1999).
- Ou, S. H. *Rice Diseases* 2nd edn (Commonwealth Mycological Institute, Kew, 1985).
- Polis, G. A., Anderson, W. B. & Holt, R. D. Toward an integration of landscape and food web ecology: The dynamics of spatially subsidized food webs. *Annu. Rev. Ecol. Syst.* **28**, 289–316 (1997).
- Dwyer, G., Elkinton, J. S. & Hajek, A. E. Spatial scale and the spread of a fungal pathogen of gypsy moth. *Am. Nat.* **152**, 485–494 (1998).
- Schwartz, M. W. Choosing the appropriate scale of reserves for conservation. *Annu. Rev. Ecol. Syst.* **30**, 83–108 (1999).
- Waide, R. B. *et al.* The relationship between productivity and species richness. *Annu. Rev. Ecol. Syst.* **30**, 257–300 (1999).
- Matson, P. A., Parton, W. J., Power, A. G. & Swift, M. J. Agricultural intensification and ecosystem properties. *Science* **277**, 504–508 (1997).
- Tilman, D. The greening of the green revolution. *Nature* **396**, 211–212 (1998).
- International Rice Research Institute. *IRRI Rice Facts* (International Rice Research Institute, Manila, 1997).
- Baker, B., Zambryski, P., Staskawicz, D. & Dinesh-Kumar, S. P. Signaling in plant-microbe interactions. *Science* **276**, 726–733 (1997).
- Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G. & Jones, J. D. G. Molecular genetics of plant disease resistance. *Science* **268**, 661–667 (1995).
- Ou, S. H. Pathogenicity and host plant resistance in rice blast disease. *Annu. Rev. Phytopathol.* **18**, 167–187 (1980).
- Bonman, J. M., Khush, G. S. & Nelson, R. J. Breeding rice for resistance to pests. *Annu. Rev. Phytopathol.* **30**, 507–528 (1992).
- Kiyosawa, S. Genetics and epidemiological modeling of breakdown of plant disease resistance. *Annu. Rev. Phytopathol.* **20**, 93–117 (1982).
- Boudreau, M. A. & Mundt, C. C. in *Environmentally Safe Approaches to Disease Control* (eds Rechcigl, J. & Rechcigl, N.) 33–62 (CRC, Boca Raton, 1997).
- Chin, K. M. & Wolfe, M. S. Selection on *Erysiphe graminis* in pure and mixed stands of barley. *Plant Pathol.* **33**, 89–100 (1984).
- Calonnet, A., Goyeau, H. & de Vallavieille-Pope, C. Effects of induced resistance on infection efficiency and sporulation of *Puccinia striiformis* on seedlings in varietal mixtures and on field epidemics in pure stands. *Eur. J. Plant Pathol.* **102**, 733–741 (1996).
- George, M. L., Nelson, R. J., Zeigler, R. S. & Leung, H. Rapid population analysis of *Magnaporthe grisea* by using rep-PCR and endogenous repetitive DNA sequences. *Phytopathology* **88**, 223–229 (1998).
- Francis, C. A. in *Multiple Cropping Systems* (ed. Francis, C. A.) 1–19 (MacMillan, New York, 1986).
- Mundt, C. C. in *Plant Disease Epidemiology* (eds Leonard, K. J. & Fry, W. E.) 150–181 (McGraw-Hill, 1989).
- Mundt, C. C. & Browning, J. A. in *The Cereal Rusts* Vol. 2 (eds Roelfs, A. P. & Bushnell, W. R.) 527–560 (Academic, Orlando, 1985).
- Browning, J. A. & Frey, K. J. in *Strategies for the Control of Cereal Disease* (eds Jenkyn, J. F. & Plumb, R. T.) 37–46 (Blackwell, Oxford, 1981).
- Wolfe, M. S. in *Barley Genetics VI* (ed. Munck, L.) 1055–1067 (Munksgaard, Copenhagen, 1992).
- Mann, C. Reseeding the Green Revolution. *Science* **277**, 1038–1042 (1997).
- Schmidt, R. A. in *Plant Disease: An Advanced Treatise* (eds Horsfall, J. G. & Cowling, E. B.) 287–315 (Academic, New York, 1978).

29. *The State Standard of the People's Republic of China*, No. GB/T 15790-1995. 1–13 (China Standard Press, Beijing, 1996).

Acknowledgements

This work was supported by the Asian Development Bank, the Yunnan Province Government, The Ministry of Science and Technology of China, the International Rice Research Institute (IRRI), and a scientific agreement between IRRI and Oregon State University. We thank the personnel of the provincial and county Plant Protection Stations and participating farmers for their contributions to this project, and M. Hoffer for computer assistance and graphics.

Correspondence and requests for materials should be addressed to C.C.M. (e-mail: mundtc@bcc.orst.edu).

Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval

Karim Nader, Glenn E. Schafe & Joseph E. Le Doux

W. M. Keck Foundation Laboratory of Neurobiology, Center for Neural Science, New York University, New York, New York 10003, USA

'New' memories are initially labile and sensitive to disruption before being consolidated into stable long-term memories^{1–5}. Much evidence indicates that this consolidation involves the synthesis of new proteins in neurons^{6–9}. The lateral and basal nuclei of the amygdala (LBA) are believed to be a site of memory storage in fear learning¹⁰. Infusion of the protein synthesis inhibitor anisomycin into the LBA shortly after training prevents

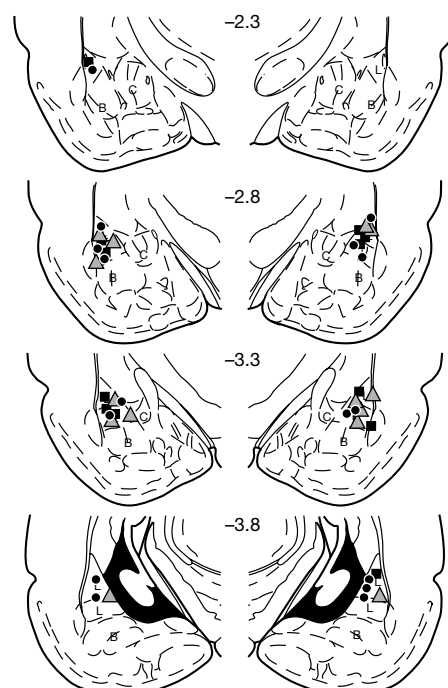


Figure 1 Schematic representation of the amygdala at four different rostral-caudal planes. The numbers represent the posterior coordinate from bregma. Injector placements in the LBA are represented by the filled symbols; black filled squares represent ASCF group placements, grey filled triangles represent the low-dose anisomycin, and black filled circles represent high-dose group. L, lateral nucleus; B, basal nucleus; C, central nucleus. The placements for subsequent experiments all demonstrate similar distributions as in this experiment and therefore are not shown.

consolidation of fear memories¹¹. Here we show that consolidated fear memories, when reactivated during retrieval, return to a labile state in which infusion of anisomycin shortly after memory reactivation produces amnesia on later tests, regardless of whether reactivation was performed 1 or 14 days after conditioning. The same treatment with anisomycin, in the absence of memory reactivation, left memory intact. Consistent with a time-limited role for protein synthesis production in consolidation, delay of the infusion until six hours after memory reactivation produced no amnesia. Our data show that consolidated fear memories, when reactivated, return to a labile state that requires *de novo* protein synthesis for reconsolidation. These findings are not predicted by traditional theories of memory consolidation.

The idea that new memories go through an initial labile period before being consolidated into stable long-term memories is an entrenched part of psychological and neurobiological models of memory¹². For example, there is considerable evidence that the formation of a long-term memory can be disrupted by certain treatments, such as systemic drug injections or electroconvulsive shock, given shortly after training, but that the same treatments given several hours or days later have no effect. One of the most commonly used drug manipulations involves the administration of drugs that block the translation of RNA into protein. Studies of this type indicate that memory consolidation involves protein synthesis^{5–9}.

It has also been reported that electroconvulsive shock or systemic drug administration given after memory reactivation (retrieval) can cause an amnesia for the original learning^{13–16}, which indicates that consolidated memories might become labile when retrieved, and might even require reconsolidation. Here we examine whether reconsolidation involving protein synthesis is required for retrieved memories to persist. We use a behavioural paradigm, auditory fear conditioning, for which the neural circuit underlying memory formation is well characterized^{17–19}. This allows us to manipulate

memory at its presumed locus of storage, in contrast to past studies in which drugs were administered systemically. Specifically, we target infusions of anisomycin, an inhibitor of protein synthesis, to the LBA, a region implicated in fear learning by lesion, pharmacological and physiological findings^{17–19}. Previously, we showed that infusions of anisomycin given after training block long-term but not short-term memory of auditory fear conditioning¹¹. Here we examine the effects of similar manipulations administered after retrieval.

Rats were given a single pairing of a tone (conditioned stimulus, CS) and foot-shock (unconditioned stimulus, US). On test days, immobility (freezing) was used as an index of fear learning²⁰. Twenty-four hours later, the rats received a single CS presentation (test 1) immediately followed by bilateral infusions of anisomycin or vehicle (artificial cerebrospinal fluid; ACSF) into the LBA. Freezing in test 1 was specific to the CS and comparable across groups. An analysis of variance (ANOVA) that compared freezing during the pre-CS or CS periods across groups indicated that there was no interaction between these two variables ($F(2, 18) = 1.6$), nor an effect of group ($F(2, 18) = 1.4$). However, there was a significant effect of period ($F(1, 18) = 160, P < 0.01$). Twenty-four hours after test 1, the rats were presented with three CSs (test 2). Anisomycin produced a dose-dependent decrease in freezing in response to the CS in test 2 (Figs 1, 2a–c). An ANOVA revealed a main effect of group ($F(2, 18) = 12, P < 0.01$). A Newman–Keuls post hoc test revealed that the low-dose anisomycin and ACSF groups were similar to each other ($P > 0.05$), but both were significantly different from the high-dose group (P values < 0.01). Extinction was observed over the three CS presentations (main effect of trial ($F(2, 36) = 7, P < 0.01$), but there was no interaction between trials and group ($F < 1$). This effect of anisomycin requires that the memory be actively retrieved, as omission of the CS before anisomycin infusion in test 1 led to normal conditioned fear responses in test 2 (Fig. 2d, e; no main effect of group ($F < 1$). The latter finding

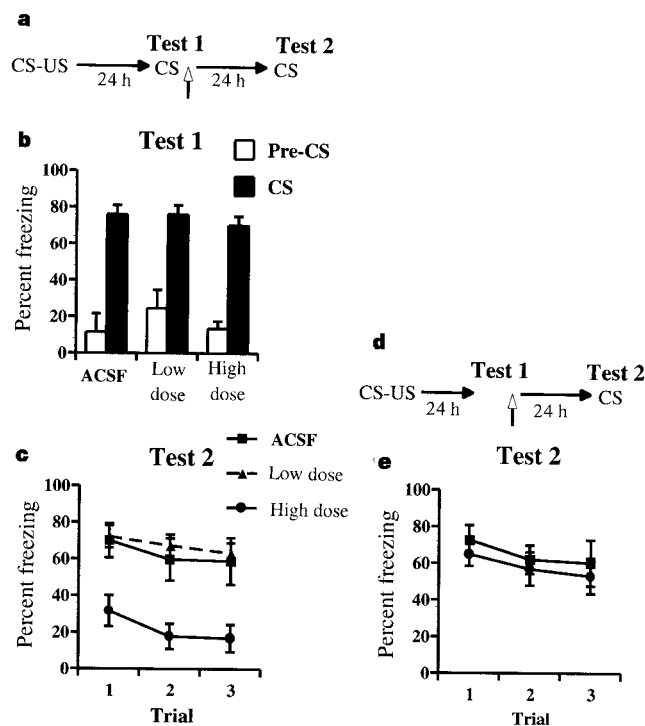


Figure 2 A test of whether consolidated fear memories can become labile when reactivated. **a**, The behavioural procedure used for experiment 1A. **b**, Freezing to the CS on test 1 was comparable across groups and was specific to the CS. **c**, Intra-LBA anisomycin infusions after reactivation of a consolidated fear memory produce amnesia for the original learning, as measured on test 2. **d**, **e**, Rats demonstrated normal memory if

the CS was omitted before anisomycin. **d**, The behavioural procedure used for experiment 1B. Rats were placed in the test chamber and received infusions of anisomycin. **e**, Percent freezing on test 2. Figure legend is applicable to both **c** and **e**. Vertical open-headed arrows represent infusions. All data points represent group means $\pm$ s.e.m.

rules out the possibility that the deficit we observed is due to a disruption of late waves of protein synthesis²¹ that may be necessary for the consolidation of the original learning, or to nonspecific effects such as damage to the amygdala by the drug. It is unlikely that the high dose of anisomycin damaged the amygdala after CS transmission, as there was no histological evidence of amygdala damage. Furthermore, rats relearn fear conditioning normally when the same dose of anisomycin is infused into the amygdala after initial learning. The fact that anisomycin infusions after reactivation of the fear memory produced amnesia for the original learning indicates that reactivation of a consolidation fear memory may place it in a labile state, one that has to be reconsolidated by protein synthesis to remain usable to the organism in future situations.

Protein synthesis inhibitors typically impair the consolidation of new memories when they are administered during a specific time window (which varies from minutes to hours) after learning⁸. Administration of such drugs after this time window does not affect memory. We next asked whether a time window also exists for reconsolidation, by delaying anisomycin infusions for 6 h after retrieval. Freezing in test 1 was specific to the CS and comparable across groups. ANOVAs revealed no period $\times$ group interaction ($F < 1$) and no effect of group ($F = 1$) but a significant effect of period ($F(1, 13) = 124, P < 0.01$). In contrast to anisomycin infusion immediately after retrieval, infusion 6 h after retrieval had no effect (Fig. 3). ANOVAs revealed that in test 2 there was no main effect of group ($F(1, 13) = 4$), no trial $\times$ group interaction ($F(2, 26) = 2.7$) and no main effect of trial ($F(2, 26) = 3$). Note that the nonsignificant impairment produced by delaying anisomycin is much smaller than that seen when anisomycin is given immediately after CS reactivation (compare with Fig. 2c). Thus, both consolidation and reconsolidation have time windows within which protein synthesis is required if a memory is to persist.

In all of the previous experiments, the time between training and CS presentation for test 1 was about 24 h. We next investigated whether reconsolidation has a temporal gradient. For example, older memories may be more thoroughly consolidated and thus may be resistant to becoming unstable when retrieved. To test this, we waited for 14 days between conditioning and test 1 (Fig. 4).

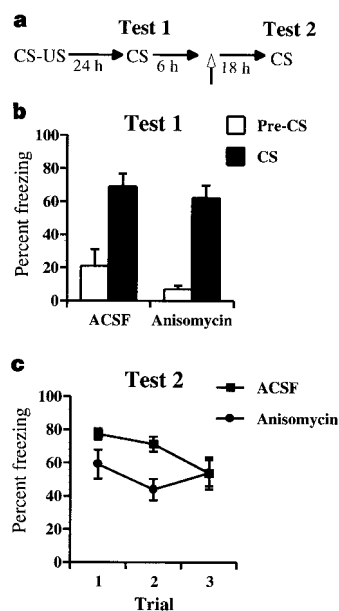


Figure 3 Intact memory if anisomycin infusions are delayed by 6 h. **a**, The behavioural procedure used for experiment 2. Vertical open-headed arrows represent infusions. **b**, Freezing on test 1 was specific to the CS and comparable across groups. **c**, Percent freezing during test 2. The groups are not significantly different. All data points represent group means $\pm$ s.e.m.

Postponing test 1 by 14 days produced an incubation effect²² such that freezing in test 1 was higher than in the previous experiment. Freezing during test 1 was specific to the CS and was comparable across groups. There was no period $\times$ group interaction ($F(1, 9) = 1.1$) and no effect of group ($F = 1$), but a significant effect of period ($F(1, 9) = 440, P < 0.01$). However, the performance of the groups in test 2 diverged—the group receiving intra-LBA infusions of anisomycin showed significantly less freezing than the controls. ANOVAs revealed a main effect of group ($F(1, 9) = 20.6, P < 0.01$) and trials ($F(2, 18) = 7, P < 0.01$), but no interaction between these variables ($F < 1$). Thus, blockade of protein synthesis in the LBA after memory reactivation caused amnesia of the original learning, even though the learning took place 14 days before the reactivation and drug treatment. Even well consolidated memories are labile and subject to disruption when reactivated.

Although anisomycin blocks reconsolidation, it is possible that the impairment is due to nonspecific effects that render the amygdala temporarily dysfunctional for reasons other than protein synthesis inhibition. To provide compelling evidence that any manipulation acts specifically on the molecular mechanisms mediating consolidation, as opposed to producing nonspecific effects, at a minimum it is necessary to show that memory is intact shortly after training but impaired later^{23–26}. For example, rats freeze normally 4 h but are impaired 24 h after receiving intra-LBA anisomycin infusions immediately after training¹¹. Using the same logic and applying it to reconsolidation, if the effects of anisomycin infused into the LBA shortly after memory reactivation are specific to reconsolidation, then freezing should be normal at 4 h, but impaired at 24 h, after CS presentation. We refer to these two time points as post-reactivation short-term memory (PR-STM) and post-reactivation long-term memory (PR-LTM), respectively.

The two groups exhibited comparable freezing during test 1

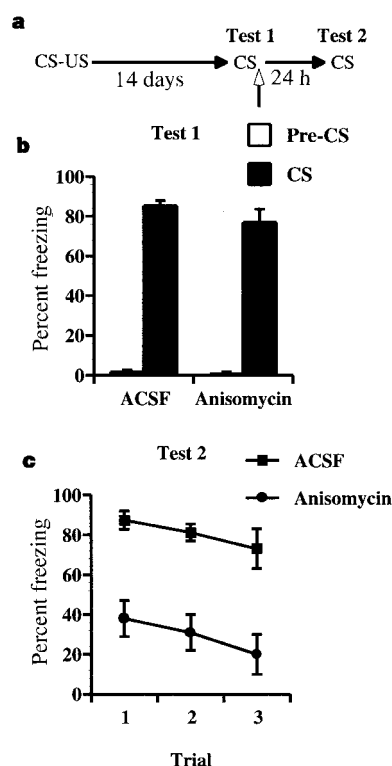


Figure 4 Fourteen days after training, anisomycin infusions after reactivation of the memory still produce amnesia. **a**, The behavioural procedure used for experiment 3. Vertical open-headed arrows represent infusions. **b**, Freezing during test 1 was specific to the CS and was comparable across groups. **c**, Percent freezing on test 2. All data points represent group means $\pm$ s.e.m.

(Fig. 5). An ANOVA revealed a significant effect of period ($F(1, 10) = 147$, $P < 0.01$), no effect of group ($F < 1$) and no interaction between these two variables ($F < 1$). Rats treated with anisomycin immediately after test 1 exhibited intact PR-STM but deficient PR-LTM. ANOVAs on the scores of the PR-STM test showed that there was no main effect of group ($F < 1$), no interaction between trial $\times$ group ($F < 1$) and no effect of trial ($F < 1$). Conversely, a similar analysis on the PR-LTM scores did reveal a significant effect of group ($F(1, 14) = 14$, $P < 0.01$). Furthermore, there was no interaction between group and trial ($F < 1$) but there was a main effect of trial ($F(2, 28) = 5$, $P < 0.05$). The fact that animals can accurately perceive, evaluate and respond to the CS 4 h after anisomycin infusion shows that the amygdala is functionally intact at the time of the PR-STM test, and thus that anisomycin did not affect reconsolidation by producing nonspecific effects. This pattern of findings, intact PR-STM and impaired PR-LTM, localizes the effects of anisomycin on fear behaviour to the molecular processes mediating reconsolidation.

The conventional view of memory consolidation predicts that blockade of protein synthesis should block new learning in test 1 of these experiments, which is an extinction test. The only new learning that occurs is about the failure of the CS to predict the US. Blockade of protein synthesis should therefore block extinction, and thus enhance memory, according to the conventional view. In contrast, however, blockade of protein synthesis had the opposite effect—it eliminated the memory rather than making it stronger.

These results provide evidence that fear memories, once retrieved, must undergo protein-synthesis-dependent reconsolidation in the LBA or nearby areas to remain accessible at later times. The full implications of this finding for fear and other memories are not understood at present. It is possible that not all memories require reconsolidation. There may be a range of parameters within which reactivation of a memory converts it into a labile state, possibly involving the extent of experience with the particular

learning situation, the kind of learning system engaged and the motivational state of the subject at the time of learning and retrieval.

Reconsolidation may reflect the dynamic nature of the process by which new information is added to existing stores. It has long been believed that memory retrieval is an active or constructive process by which old information is integrated with the current knowledge base of the organism²⁷. Reconsolidation may be part of the neural mechanism through which constructed memories are stored for later constructions.

Current models of learning propose that the production of new proteins is necessary for structural encoding of recent experiences in long-term memory^{6–9}. In addition, it now appears that new proteins are also required to maintain memories that have been reactivated. It seems unlikely that retrieval reverses the structural changes induced by original learning. Rather, some property of retrieval may destabilize the structural changes such that they now have to be reconsolidated with the aid of new proteins. The fact that animals demonstrate intact freezing 4 h after anisomycin indicates that the structural changes may remain functional for at least 4 h. Particularly important for future work will be the clarification of the physiological basis of memory lability during retrieval and the requirement for reconsolidation. It is possible that some modification of the synaptic tagging hypothesis^{28,29}, which proposes that active synapses are given molecular markers that help stabilize synapses by capturing proteins made in the cell nucleus, might account for lability and reconsolidation, although this remains to be seen.

We have shown that reactivation changes the status of a consolidated fear memory to a labile one that must be reconsolidated using de novo protein synthesis to persist. Like consolidation itself, reconsolidation has a temporal window during which blockade of protein synthesis is effective, and beyond which it is not. Furthermore the reconsolidation process, like the consolidation process, has a short-term phase that is not dependent on protein synthesis. A definition of consolidation based on 'new' memories is insufficiently broad to describe these data. We propose, in keeping with the original suggestion by Misanin *et al.*¹³, that as a first approximation 'active' rather than 'new' memories be viewed as labile, subject to disruption, and requiring protein-synthesis-dependent consolidation processes. □

Methods

Subjects

Subjects were adult male Sprague–Dawley rats from Hilltop Labs. Rats were housed individually in plastic Nalgene cages and maintained on a 12/12 h light/dark cycle. Food and water were provided *ad libitum*.

Surgery and histology

Under Nembutal anaesthesia (45 mg kg⁻¹), rats were implanted bilaterally with 22-gauge stainless steel cannulas into the lateral amygdala. Coordinates were 3.0 mm posterior to bregma, 5.3 mm lateral to the midline and 8.0 mm ventral to the skull surface. Rats were given at least 5 days to recover before experimental procedures. All procedures were in accordance with the NIH Guide, and were approved by the NYU Animal Care and Use Committee. At the end of the experiment, using standard histological methods, animals were perfused and their brains sectioned at 50 μ m thickness. The sections were stained using cresyl violet and examined by light microscopy for cannula penetration into the LBA.

Intra-LBA infusions

Drugs were infused slowly using an infusion pump into the LBA at 0.25 μ l min⁻¹. Following drug infusion, cannulas were left in place for an additional minute to allow diffusion of the drug away from the cannula tip. Anisomycin (Sigma) was dissolved in equimolar HCl, diluted with ACSF and adjusted to pH 7.4 with NaOH. Although the lateral nucleus was the main target, the 0.5- μ l infusions also probably affected the adjacent basal nucleus. We therefore refer to the affected area as the lateral and basal amygdala (LBA).

Apparatus

Conditioning and tone testing were conducted in different chambers. For conditioning, rats were placed in a Plexiglas rodent conditioning chamber (chamber A) with a metal grid floor (Model E10-10, Coulbourn Instruments) that was enclosed within a sound attenuating chamber (Model E10-20). The chamber was dimly illuminated by a single house light. For tone testing, rats were placed in a different Plexiglas chamber (ENV-001,

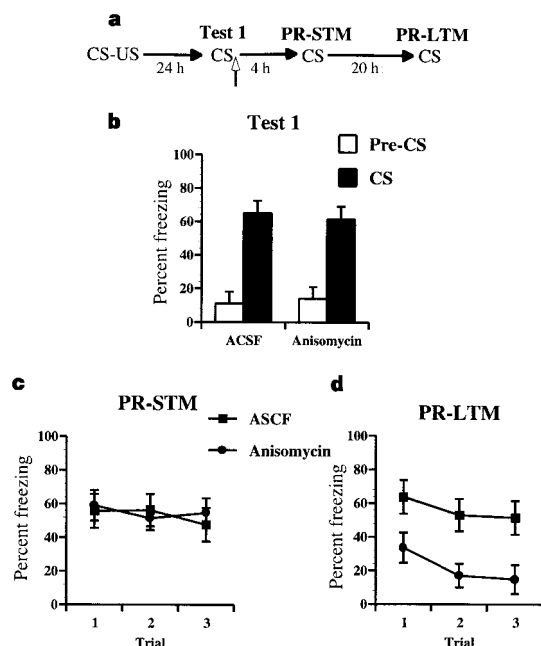


Figure 5 Amnesia following anisomycin is not due to nonspecific effects. **a**, The behavioural procedure used for experiment 4. Vertical open-headed arrows represent infusions. Anisomycin infusions impaired post-reactivation long-term memory (PR-LTM), but had no effect on post-reactivation short-term memory (PR-STM). **b**, All rats demonstrated comparable freezing scores on test 1. **c**, Scores on the PR-STM test 4 h after reactivation and anisomycin. **d**, Scores on the PR-LTM test 24 h after reactivation and anisomycin. The key is applicable to both **c** and **d**. All data points represent group means $\pm$ s.e.m.

MedAssociates), which has been shown to minimize generalization from the conditioning environment¹¹. The tone-testing chamber (chamber B) was brightly lit with three house lights and contained a flat black Formica floor that had been washed with peppermint soap. A micro-video camera was mounted at the top of the chamber so that rats could be videotaped during testing.

General behavioural procedures

Rats were placed in chamber A and after a 5-min acclimatizing period, given a single conditioning trial consisting of a 30-s presentation of a 5-kHz, 75-dB tone CS that ended at the same time as a 2.0-mA, 1-s food shock US. Rats were then returned to their home cages. The next day, 24 h later, rats were placed in chamber B and given a single 30-s CS presentation (test 1) to reactivate the memory. Twenty-four hours after test 1, rats were returned to chamber B and given three CS presentations (test 2).

Experiment 1A

Rats were infused with either 62.5 µg per 0.5 µl per side ($n = 8$) or 6.2 µg per 0.5 µl per side ($n = 7$) anisomycin or ACSF ($n = 6$) immediately after CS termination during test 1. The highest dose of anisomycin was chosen based on a previous study showing >90% suppression of protein synthesis in cortex using this concentration³⁰. Previous data have shown that post-training intra-LBA infusions of the high but not low dose blocked consolidation of fear conditioning¹¹.

Experiment 1B

During test 1 no CS was presented while the animals explored chamber B, but rats still received an infusion of vehicle ($n = 6$) or the high dose of anisomycin ($n = 7$) at the end of the exposure to chamber B.

Experiment 2

High-dose anisomycin ($n = 8$) or vehicle ($n = 7$) infusions were performed 6 h after test 1. Animals were transported to the infusion room, received the infusion and were then returned to their home cage.

Experiment 3

Fourteen days were inserted between conditioning and test 1. After CS reactivation, rats received either high-dose anisomycin ($n = 6$) or vehicle ($n = 5$) infusion.

Experiment 4

After test 1 animals received either vehicle ($n = 8$) or high-dose anisomycin ($n = 8$) infusions into the LBA. An extra test was inserted 4 h after test 1 (post-reactivation short-term memory, PR-STM) during which animals received three CS presentations. Test 2 was performed as described above, 24 h after reactivation (post-reactivation long-term memory, PR-LTM). These time points were chosen based on the findings that freezing is intact 4 h, but impaired 24 h, after conditioning¹¹.

Received 3 April; accepted 18 May 2000.

- Müller, G. E. & Pilzecker, A. Experimentelle beiträge zur lehre vom gedächtnis. *Z. Psychol.* (Suppl. 1) (1900).
- Burnham, W. H. Retroactive amnesia: illustrative cases and a tentative explanation. *Am. J. Psychol.* **14**, 382–396 (1903).
- Duncan, C. P. The retroactive effect of electroconvulsive shock. *J. Comp. Physiol. Psychol.* **42**, 32–44 (1949).
- Hebb, D. O. *The Organization of Behavior* (Wiley, New York, 1949).
- Dudai, Y. Consolidation: Fragility on the road to the engram. *Neuron* **17**, 367–370 (1996).
- Goel, P., Castellucci, V. F., Schacher, S. & Kandel, E. R. The long and short of long-term memory—a molecular framework. *Nature* **322**, 419–422 (1986).
- Flexner, L. B., Flexner, J. B. & Stellar, E. Memory and cerebral protein synthesis in mice as affected by graded amounts of puromycin. *Exp. Neurol.* **13**, 264–272 (1965).
- Davis, H. P. & Squire, L. R. Protein synthesis and memory: A review. *Psychol. Bull.* **96**, 518–559 (1984).
- Agranoff, B. W. in *Basic Neurochemistry* (eds Siegel, G. J., Albers, R. W., Agranoff, B. W. & Catzman, R.) 801–820 (Little, Brown, Boston, 1981).
- Fanselow, M. S. & LeDoux, J. E. Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* **23**, 229–232 (1999).
- Schafe, G. E. & LeDoux, J. E. Memory consolidation of auditory Pavlovian fear conditioning requires protein synthesis and PKA in the amygdala. *J. Neurosci.* (in the press).
- McGaugh, J. L. Memory—a century of consolidation. *Science* **287**, 248–251 (2000).
- Misanin, J. R., Miller, R. R. & Lewis, D. J. Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science* **160**, 203–204 (1968).
- Judge, M. E. & Quartermain, D. Characteristics of retrograde amnesia following reactivation of memory in mice. *Physiol. Behav.* **28**, 585–590 (1982).
- Sara, S. J. Retrieval and reconsolidation: toward a neurobiology of remembering. *Learn. Mem.* **7**, 73–84 (2000).
- Mactutus, C. F., Riccio, D. C. & Ferek, J. M. Retrograde amnesia for old (reactivated) memory: Some anomalous characteristics. *Science* **204**, 1319–1320 (1979).
- Davis, M. Neurobiology of fear responses: the role of the amygdala. *J. Neuropsychiat. Clin. Neurosci.* **9**, 382–402 (1997).
- Fanselow, M. S. Pavlovian conditioning, negative feedback, and blocking: mechanisms that regulate association formation. *Neuron* **20**, 625–627 (1998).
- LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
- Blanchard, R. J. & Blanchard, D. C. Crouching as an index of fear. *J. Comp. Physiol. Psychol.* **67**, 370–375 (1969).

- Scholey, A. B., Rose, S. P., Zamani, M. R., Bock, E. & Schachner, M. A role for the neural cell adhesion molecule in a late, consolidating phase of glycoprotein synthesis six hours following passive avoidance training of the young chick. *Neuroscience* **55**, 499–509 (1993).
- Brady, J. V. The effects of electroconvulsive shock on a conditioned emotional response: the permanence effect. *J. Comp. Physiol. Psychol.* **45**, 9–13 (1952).
- Abel, T. *et al.* Genetic demonstration of a role for PKA in the late phase of LTP and in hippocampus-based long-term memory. *Cell* **88**, 615–626 (1997).
- Bourtchouladze, R. *et al.* Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79**, 59–68 (1994).
- Guzowski, J. F. & McGaugh, J. L. Antisense oligodeoxynucleotide-mediated disruption of hippocampal cAMP response element binding protein levels impairs consolidation of memory for water maze training. *Proc. Natl Acad. Sci. USA* **94**, 2693–2698 (1997).
- Lamprecht, R., Harzvi, S. & Dudai, Y. cAMP response element-binding protein in the amygdala is required for long- but not short-term conditioned taste aversion memory. *J. Neurosci.* **17**, 8443–8450 (1997).
- Bartlett, F. C. *Remembering* (Cambridge Univ. Press, Cambridge, 1932).
- Martin, K. C. *et al.* Synapse-specific, long-term facilitation of aplysia sensory to motor synapses: a function for local protein synthesis in memory storage. *Cell* **91**, 927–938 (1997).
- Frey, U. & Morris, R. G. Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536 (1997).
- Rosenblum, K., Meiri, N. & Dudai, Y. Taste memory: the role of protein synthesis in gustatory cortex. *Behav. Neural Biol.* **59**, 49–56 (1993).

Acknowledgements

This research was supported in part by NIMH grants to J.E.L. and a HFSF grant to K.N. The work was also supported by a grant from the W. M. Keck Foundation to N.Y.U. The authors thank A. Schoute for technical assistance.

Correspondence and requests for materials should be addressed to K.N. (e-mail: karim@cns.nyu.edu).

Cortex-restricted disruption of NMDAR1 impairs neuronal patterns in the barrel cortex

Takuji Iwasato*, Akash Datwani†, Alexander M. Wolf‡, Hiroshi Nishiyama*, Yusuke Taguchi*, Susumu Tonegawa||, Thomas Knöpfel‡, Reha S. Erzurumlu† & Shigeyoshi Itohara*

*Laboratory for Behavioral Genetics, Brain Science Institute (BSI), RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

†Department of Cell Biology and Anatomy and Neuroscience Center, LSUHSC, New Orleans, Louisiana 70112, USA

‡Laboratory for Neuronal Circuit Dynamics, Brain Science Institute (BSI), RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

§Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan
||Howard Hughes Medical Institute, Center for Learning & Memory, RIKEN-MIT Neuroscience Research Center, Department of Biology, MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

In the rodent primary somatosensory cortex, the configuration of whiskers and sinus hairs on the snout and of receptor-dense zones on the paws is topographically represented as discrete modules of layer IV granule cells (barrels) and thalamocortical afferent terminals^{1,2}. The role of neural activity, particularly activity mediated by NMDARs (N-methyl-D-aspartate receptors), in patterning of the somatosensory cortex has been a subject of debate^{3–6}. We have generated mice in which deletion of the *NMDAR1* (*NR1*) gene is restricted to excitatory cortical neurons, and here we show that sensory periphery-related patterns develop normally in the brainstem and thalamic somatosensory relay stations of these mice. In the somatosensory cortex, thalamocortical afferents corresponding to large whiskers form patterns and display critical period plasticity, but their patterning is not as distinct as that seen in the cortex of normal mice. Other thalamocortical patterns corresponding to sinus hairs and digits are mostly absent. The cellular aggregates known as barrels and barrel boundaries do not develop even at sites where thalamocortical afferents cluster. Our findings indicate that cortical NMDARs

are essential for the aggregation of layer IV cells into barrels and for development of the full complement of thalamocortical patterns.

The rodent somatosensory system is an excellent model to study molecular mechanisms underlying the establishment of patterned topographic connections between the sensory periphery and the brain. Trigeminal and dorsal column pathways convey a somatosensory map from the periphery to the neocortex by relay stations in the brainstem and the ventrobasal thalamus. At each level, pre-synaptic afferents and their target cells establish a patterned array of modules corresponding to the arrangement of whiskers and sinus hairs on the snout (trigeminal pathway), and receptor-dense zones of the paws (dorsal column pathway)^{2,7,8}. These patterns are consolidated during the first postnatal week, and are subject to alterations if the sensory periphery is perturbed by nerve or whisker lesions during a critical period^{2,7,8}. Thus, early neural activity might be essential in the development and plasticity of central patterns much like that of the vertebrate visual system^{9,10}. Within this context, NMDAR activity has attracted much attention^{9,11}. Initial studies using pharmacological blockade of NMDARs in postnatal rat somatosensory cortex did not detect effects on thalamocortical patterning, but revealed functional impairment of thalamocortical connectivity and whisker-specific responsiveness of barrel neurons^{4,5}. In contrast, genetic manipulations of NMDARs showed an absence of periphery-related patterns in the somatosensory cortex⁶. This result could not be ascribed to direct effects on the neocortex, however, as it could have been a consequence of impaired patterns at the brainstem level^{6,12,13}. To delineate the specific role(s) of cortical NMDARs in patterning, here we disrupted the function of NR1—the essential subunit of the NMDAR—specifically in the cortex, using the Cre/loxP system¹⁴.

We took advantage of the promoter for the homeobox gene, *Emx1*, which is expressed exclusively in the dorsal telencephalon from embryonic stages to adulthood¹⁵. We inserted the Cre recombinase gene into the *Emx1* locus by homologous recombination in

embryonic stem (ES) cells (Fig. 1a). Heterozygous *Emx1*-Cre (*Emx1*^{Cre/+}) mice were intercrossed to obtain homozygous (*Emx1*^{Cre/Cre}) mice (Fig. 1b). The *Emx1*^{Cre/Cre} mice are viable and fertile, as are the homozygous *Emx1* knockout mice¹⁶. To assess the temporal and spatial specificity of Cre-mediated recombination in *Emx1*-Cre mice, we crossed *Emx1*^{Cre/Cre} mice with CAG-CAT-Z reporter mice¹⁷, and obtained double heterozygous (*Emx1*-Cre/LacZ) mice. *Emx1*-Cre/LacZ embryos (embryonic day (E) 11.5–12.5, *n* = 4) displayed strong X-gal staining (indicating Cre-mediated recombination) in the dorsal telencephalon (Fig. 1c). Staining of postnatal brains (postnatal day (P) 0–8, *n* = 23) indicated extensive recombination in the neocortex, hippocampus and olfactory bulb, whereas recombination in other brain regions was negligible (Fig. 1d). The barrels are consolidated during the first postnatal week through inputs from subcortical somatosensory nuclei^{8,18}. We detected very little or no Cre-mediated recombination in these nuclei during this period or earlier (see Supplementary Information).

We crossed the *Emx1*^{Cre/Cre} mice with heterozygous *NR1* null mutant (*NR1*^{+/-}) mice¹² to obtain double heterozygous (*Emx1*^{Cre/+} *NR1*^{+/-}) mice, and further crossed these with homozygous floxed *NR1* (*NR1*^{lox/lox}) mice¹⁴ to obtain four types of mice: *Emx1*^{Cre/+} *NR1*^{lox/+}; *Emx1*^{+/+} *NR1*^{lox/+}; *Emx1*^{Cre/+} *NR1*^{lox/-}; and *Emx1*^{+/+} *NR1*^{lox/-}. *Emx1*^{Cre/+} *NR1*^{lox/-} mice constitute cortex-specific *NR1* knockout mice, and hereafter we refer to them as CxNR1KO mice. The CxNR1KO mice were viable and their body weight was not significantly different from that of control littermates at P0, however, their growth rate was relatively low. At P7, the average body weight of CxNR1KO mice was about 70% of that of control littermates.

We examined the *NR1* gene disruption in the developing CxNR1KO cortex by several independent measures. Western blot analysis showed that the thalamus and brainstem of CxNR1KO mice have similar levels of NR1 protein expression to those of *Emx1*^{+/+} *NR1*^{lox/-} (floxed) control mice (Fig. 2a). In contrast, the cortex of CxNR1KO mice exhibited greatly reduced NR1 expression

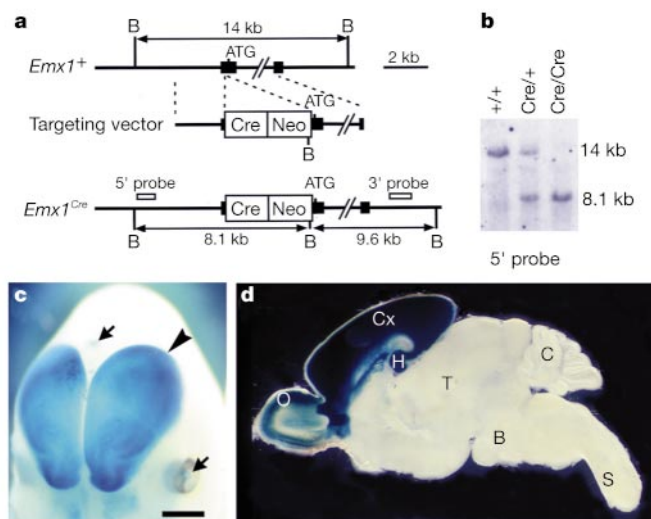


Figure 1 Generation of *Emx1*-Cre mice and their cortex-restricted recombination. **a**, Wild-type *Emx1* allele (*Emx1*⁺), the targeting vector and targeted allele (*Emx1*^{Cre}). Filled boxes represent exons. The Cre recombinase gene (Cre) and the *pgk-neo* gene (Neo) are inserted immediately before the translation initiation codon ATG. B, *Bgl*II. **b**, Southern blot analysis. Genomic DNA was digested with *Bgl*II. **c**, Partial view of whole mount X-gal staining of an E12.5 *Emx1*-Cre/LacZ embryo. Blue staining (Cre-mediated recombination) is restricted to the dorsal telencephalon (arrowhead). Arrows indicate staining in non-neural tissues. **d**, X-gal staining of a parasagittal brain section from a P7 *Emx1*-Cre/LacZ mouse. Olfactory bulb (O), hippocampus (H) and neocortex (Cx) are strongly stained; thalamus (T), brainstem (B), cerebellum (C) and spinal cord (S) are not stained. Scale bar, 800 μ m (**c**) and 1,600 μ m (**d**).

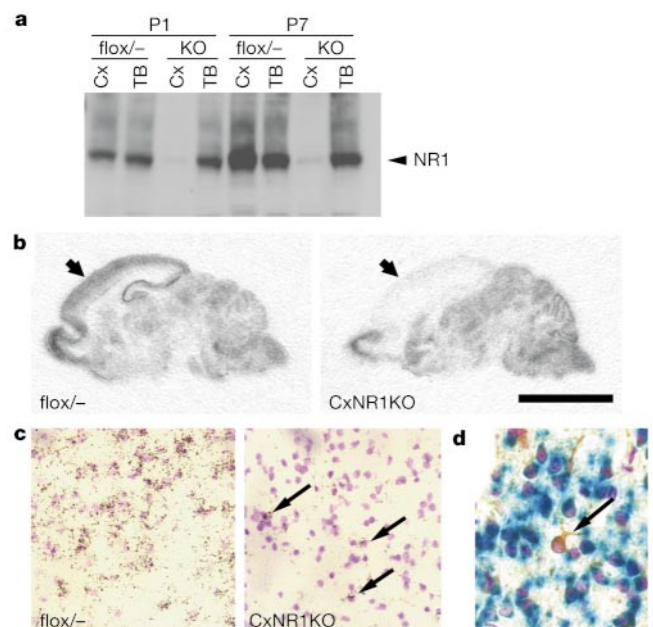


Figure 2 Cortex-restricted *NR1* disruption in CxNR1KO mice. **a**, Western blot analysis of NR1 protein expression in cortex (Cx) and thalamus and brainstem (TB) of floxed control and CxNR1KO (KO) mice at P1 and P7. **b**, *In situ* hybridization of *NR1* mRNA in sagittal sections of P7 brains. Arrows indicate cortex. **c**, *In situ* hybridization of barrel cortex layer IV at P7. CxNR1KO cortex has a few *NR1* expressing neurons (arrows). **d**, Barrel cortex layer IV of P7 *Emx1*-Cre/LacZ mouse stained with X-gal (blue), GABA antibody (brown) and cresyl violet (purple). A GABA-containing neuron (arrow) is LacZ negative. Other neurons are LacZ positive. Scale bar, 4 mm (**b**), 100 μ m (**c**) and 50 μ m (**d**).

relative to flox/– controls. At P7, the level of NR1 protein present in CxNR1KO cortex was reduced to about 5% of the level present in flox/– cortex. We then examined the distribution of NR1 messenger RNA by *in situ* hybridization and found that the amount of NR1 RNA in the neocortex and hippocampus of CxNR1KO mice was much lower than that of age-matched flox/– mice (Fig. 2b). Other brain regions exhibited similar levels of NR1 mRNA in CxNR1KO and flox/– mice. Only a few cells had detectable NR1 expression in the CxNR1KO cortex (Fig. 2c). These NR1-expressing neurons were scattered throughout the cortical layers and hippocampus, with a distribution pattern similar to that of GABA-containing interneurons¹⁹. We examined brain slices of P7 *Emx1*-Cre/LacZ mice with anti-GABA/X-gal double staining and anti-GABA/X-gal/cresyl violet triple staining (Fig. 2d). GABA-containing neurons are LacZ negative in these mice. These cortical interneurons most probably maintained their NR1 expression because they derived from the ganglionic eminence of the ventral telencephalon²⁰ where *Emx1* gene is not expressed.

We examined NMDAR-mediated activity by optical imaging using a voltage-sensitive dye in control (flox/–, *n* = 11) and

CxNR1KO (*n* = 12) cortical slices at P7 (Fig. 3). Stimulation in the white matter close to layer VI evoked postsynaptic responses in a band of overlying cortex²¹ (Fig. 3a, c). In control slices, removal of Mg²⁺ from the artificial cerebrospinal fluid (ACSF) increased this response to 350 ± 42% of that in normal ACSF 30–50 ms after stimulation (Fig. 3d). Addition of the NMDAR antagonist APV diminished this response to 70 ± 5% of the level originally observed in ACSF (Fig. 3d), confirming that NMDAR-mediated excitation occurs in normal ACSF and is enhanced by Mg²⁺ removal²². In CxNR1KO mice, Mg²⁺ removal decreased the response to 75 ± 4% of the normal ACSF level, and addition of APV reversed this reduction (Fig. 3d). These results show that the CxNR1KO barrel cortex lacks NMDAR-mediated excitation and also support the histological evidence that inhibitory neurons express NMDAR in the CxNR1KO cortex. Addition of the AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate receptor antagonist NBQX almost entirely blocked the observed responses (Fig. 3d). Collectively, the biochemical, histological and physiological data indicate that NMDAR disruption is restricted to excitatory cortical neurons in the CxNR1KO mice.

We performed a series of analyses to define the role of NMDARs in patterning of the anatomical somatosensory body map in the barrel cortex, in which we used flox/–, *Emx1*^{Cre/+}NR1^{flox/+} and NR1^{flox/+} littermates as controls. CxNR1KO mice had no abnormalities in the organization of the whiskers on the snout, or in neural patterns in the brainstem trigeminal (see Supplementary Information), dorsal column nuclei and the ventrobasal thalamus (Fig. 4b), as assessed by cytochrome oxidase histochemistry.

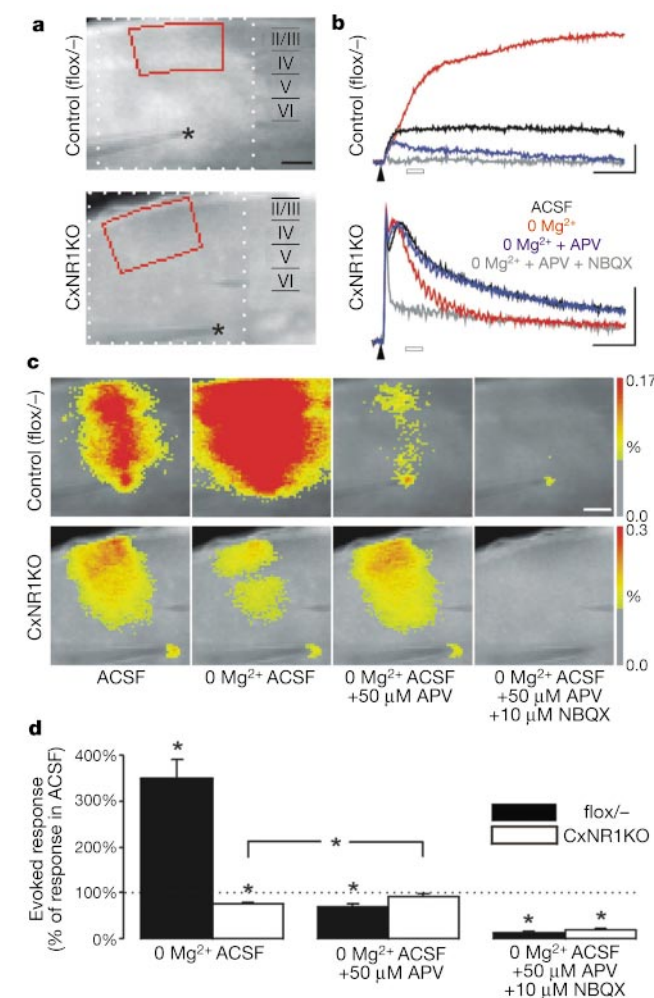


Figure 3 Lack of NMDAR-mediated excitation in barrel cortex of CxNR1KO mice. **a**, Fluorescence image of flox/– and CxNR1KO slices stained with a voltage-sensitive dye. The stimulating electrode was placed in white matter (asterisk). **b**, Fluorescence signals representing membrane depolarization were obtained from layer II–IV (marked red in **a**) in ACSF, Mg²⁺-free ACSF, Mg²⁺-free ACSF + APV (50 μM), and Mg²⁺-free ACSF + APV (50 μM) + NBQX (10 μM). Arrowheads indicate time of stimulation. Scales, 50 ms, –0.04% fluorescence change. **c**, Pseudocolour-coded maps of evoked depolarization 30–50 ms after stimulation (open bars in **b**). Scale bar, 250 μm (**a**, **c**). **d**, Statistical analysis of evoked responses (mean ± s.e.m.). Asterisk, *P* < 0.01.

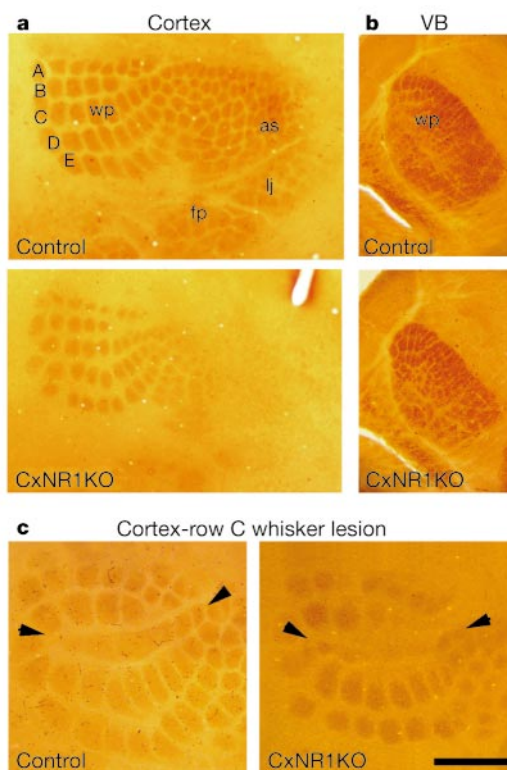


Figure 4 Whisker-related patterns as revealed by cytochrome oxidase histochemistry. **a**, Cortical patterns for five rows (A–E) of major whiskers are present but not well defined in CxNR1KO mice. **b**, Subcortical patterns in the ventrobasal thalamus (VB) are indistinguishable between control and CxNR1KO samples. **c**, Row C whisker lesions between P0 and P3 normally lead to fusion of row C barrels (arrowheads) in the cortex and expansion of neighbouring barrels (left). Similar effects are observed in CxNR1KO mice lesioned at P1 (right). Whisker pad (wp), anterior snout (as), lower jaw (lj) and forepaw (fp) representation areas are shown. Scale bar, 800 μm (**a**), 500 μm (**b**) and 450 μm (**c**).

Cytochrome-oxidase-dense patches reflect the patterned organization of both afferent terminals and target neurons²³. In contrast, in the barrel cortex there was a clear difference between the CxNR1KO and control mice at all ages studied (Fig. 4a; $n = 43$ control, $n = 68$ CxNR1KO; P5–41). In the CxNR1KO cortex, although cytochrome-oxidase-dense patches were organized into five rows in a similar orientation as in control animals, these patches were smaller and less distinct. Cytochrome oxidase patches that correspond to sinus hairs and digits were mostly absent.

We used the lipophilic tracer, DiI, (applied to ventrobasal thalamus)¹⁸ and serotonin transporter (5-HTT)-based immunohistochemistry²⁴ to visualize thalamocortical projections ($n = 15$ CxNR1KO, $n = 17$ control; both at P5). Both markers showed that thalamocortical projections displayed patches only in the large whisker representation area in the CxNR1KO mice. These patches were smaller and not as sharply defined as those seen in normal cortex (Fig. 5a, b, f). The ratio of total area of thalamocortical patches to the total area of whisker representation was significantly smaller in CxNR1KO mice ($25.08 \pm 2.78\%$; mean $\pm$ s.d., $n = 5$) than in controls ($40.92 \pm 4.24\%$; $n = 5$, $P < 0.0005$). In wild-type cortex, layer IV cells are concentrated around barrel walls, forming cell-sparse barrel hollows (centres) and septa that delineate individual barrels. In flattened and coronal cortical sections, barrels—each corresponding to a single whisker on the snout—can be easily identified with Nissl stain¹. In CxNR1KO mice, however, layer IV somatosensory cortex displayed a uniform distribution of granule cells with no indication of cellular aggregations (barrels) in all regions of the thalamocortical projection area (Fig. 5c, e, g). To further confirm the absence of barrels, we used immunohistochemistry for the extracellular matrix protein tenascin. Normally, expression of tenascin is high in the septa and defines individual barrel boundaries²⁵. At P5–7, tenascin-positive barrel boundaries were conspicuous in control animals ($n = 4$), but completely absent in CxNR1KO littermates ($n = 4$) (Fig. 5d).

Finally, we examined the effects of peripheral lesions on the patterning of whisker-related thalamocortical terminals^{4,7}. Notably, row C whisker lesions during the critical period led to fusion of row C representation areas in the CxNR1KO cortex as in control cortex, as assessed both by cytochrome oxidase histochemistry (Fig. 4c) and by 5-HTT immunohistochemistry (not shown). Thus whisker-related thalamocortical afferents can undergo structural plasticity

independent of NMDAR function in excitatory cortical cells. In lesioned CxNR1KO cases, Nissl staining showed no patterning of cortical neurons where thalamocortical terminals displayed structural plasticity (data not shown).

Our findings show that absence of NMDAR function in cortical excitatory neurons results in a lack of barrels and barrel boundaries. Thalamocortical afferents corresponding to large whiskers form periphery-related patterns and display critical period plasticity but their patterning is not well defined. Other thalamocortical patterns corresponding to sinus hairs and digits are mainly absent. Thus, cortical NMDARs are essential in the completion of thalamocortical patterning and the formation of the barrels. Studies using pharmacological blockade of NMDARs in the postnatal rat neocortex did not see any morphological abnormalities in whisker-related patterning^{4,5}, but functional alterations were pronounced⁵. Insufficient efficacy, lack of specificity and/or inappropriate timing of NMDAR blockade may have contributed to the lack of effects observed at the morphological level in these studies. In CxNR1KO mice, loss of NMDAR function begins at early embryonic stages when the cortical mantle is just forming and is restricted to excitatory cortical neurons. The morphological abnormalities observed in our CxNR1KO mice cannot be attributed to a general deficit in cortical development, because layers form normally, thalamocortical axons terminate in their proper cortical layers and somatotopy is established in the CxNR1KO mice. Furthermore, cortical cell migration is not impaired in mice with complete deletion of the *NR1* gene²⁶.

When NMDAR function is insufficient or absent in the entire brain, neuronal patterns do not form in the brainstem trigeminal nuclei^{6,12,13}. In these cases, trigeminal afferent terminals convey a topographic projection to the brainstem but fail to develop patterns, showing that NMDAR function in the brainstem is required for presynaptic patterning (a presynaptic effect). Partial and rudimentary patterning of thalamocortical axons in the somatosensory cortex of the CxNR1KO mice suggests that similar presynaptic mechanisms are also in operation in the neocortex. These results are consistent with the hypothesis that NMDARs act as 'coincidence' detectors in postsynaptic cells to consolidate coordinated afferent inputs^{9–11}. In the absence of NMDAR expression on cortical excitatory neurons, however, thalamocortical afferents corresponding to five rows of the major whiskers still form patterns. These large

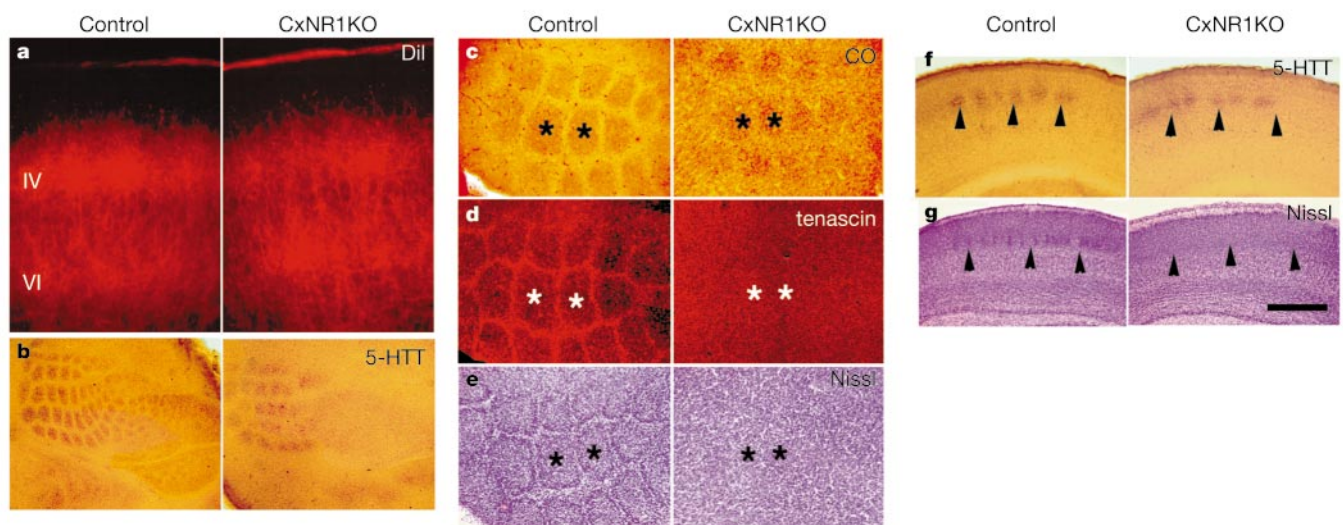


Figure 5 Partial thalamocortical axonal patterns, absence of barrels and barrel boundaries. **a**, DiI-labelled thalamocortical axons show patchy distribution in control layer IV, and a similar distribution with less distinct patches in the CxNR1KO cortex. **b**, Flattened sections stained for 5-HTT antibody also reveal distinct patterns in the control cortex and less distinct ones in the CxNR1KO cortex. In the CxNR1KO cortex, primarily the large whisker patterns are noticeable.

c–e, Cytochrome-oxidase-dense patches (asterisks) in flattened cortex (**c**), and adjacent sections immunostained for tenascin antibody (**d**). The same sections shown in **d** were counterstained with cresyl violet (**e**). **f, g**, 5-HTT antibody (**f**) or cresyl violet (**g**) stained sections spanning equivalent regions of the whisker barrel cortex in the coronal plane. Scale bar, 200 μ m (**a**), 650 μ m (**b**), 250 μ m (**c–e**) and 450 μ m (**f, g**).

whiskers have the most conspicuous neural (both axonal and cellular elements) representations along the entire somatosensory pathway. Thus, thalamocortical afferents may simply transfer the dominant patterns from the thalamus to the cortex in the absence of NMDAR function in the neocortex. This would also explain why thalamocortical afferents show critical period plasticity after row C whisker lesions. The role of remaining NMDARs in inhibitory neurons in patterning and plasticity of the thalamocortical axons remains to be determined, and might account to some extent for the patterns observed in the CxNR1KO cortex.

The complete absence of barrels, even in regions of the somatosensory cortex where some thalamocortical axons form patterns was unexpected, and reveals a new postsynaptic role for NMDARs in cortical pattern formation. This indicates that NMDARs in layer IV granule cells are crucial for detection of patterned thalamocortical inputs and construction of barrels. Barrel formation involves segregation of thalamocortical terminals in layer IV into patches and the subsequent orientation of granule cell dendrites towards these patches. Consequently, their cell bodies are displaced around axonal and dendritic clusters, and form the barrel walls²⁷. NMDARs could serve as the detectors that lead to preferential orientation and growth of dendritic trees towards thalamocortical afferent patches. Many studies in the vertebrate visual system have suggested the involvement of NMDAR in elaboration of dendritic trees (for example, see refs 28, 29), and a number of mutant phenotypes with barrel field abnormalities have been reported³⁰. Among them, knockout mice for *phospholipase C-β1* gene, which is expressed in cortex but not in thalamus, showed impairments in patterning of postsynaptic cells but not in patterning of thalamic afferents. Glutamate released from thalamic axons stimulates the phospholipase C-β1 pathway through metabotropic glutamate receptors and the NMDAR pathway, and these pathways may cooperate in the postsynaptic remodelling in the somatosensory cortex. Our results reveal the existence of pre- and postsynaptic roles of NMDAR function in cortical patterning and provide a framework for future studies to uncover the molecular mechanisms of activity-dependent refinement of sensory cortex. □

Methods

Generation of Emx1-Cre mice

A 9.4-kilobase (kb) fragment of the *Emx1* gene was used to make the targeting construct. The *NLS-Cre-poly(A)* gene and the *pgk-Neo-poly(A)* gene were inserted into a *NotI* site immediately before the *Emx1* translation initiation site, in the sense orientation. The targeting construct was transfected into ES cells (E14) using a standard protocol. Homologous recombinants were identified by Southern blotting by the presence of 8.1-kb and 9.6-kb bands in *BglII*-digested genomic DNA hybridized with 5' (*EcoRI* 0.9 kb) (Fig. 1b) and 3' (*PstI*–*EcoRI* 1 kb) (not shown) probes, respectively. The ES cell clones containing the targeting event were injected into C57BL/6 blastocysts and chimaeras were derived. Chimaeric male mice were crossed with C57BL/6 females to achieve germline transmission. All mice were maintained at the animal facilities of RIKEN-BSI according to the Institution's guidelines.

Genotyping of mice

Genotypes were determined by Southern blot and/or PCR. PCR primer sets for Emx1-Cre, CAG-CAT-Z and *NRI*^{+/+} mice were, respectively, 5'-ACCTGATGGACATGTTTCAGGG ATCG-3' and 5'-TCCGGTTATTCAACTTGCACCATGC-3'; 5'-TCGGCGGTGAAAT ATCGATGAGC-3' and 5'-CCACAGCGGATGGTTCGGATAATGC-3'; and 5'-ATGATGGGAGAGCTGCTCAG-3' and 5'-CAGACTGCCTTGGGAAAGC-3'.

Western blot

Membrane fractions were prepared, loaded (60 μg per lane) and detected with anti-NRI monoclonal antibody (Pharmingen: 54.1. 1:1,000) by standard methods. Quantification was performed with Fuji Lumino Image Analyzer LAS-1000.

In situ hybridization

In situ hybridization was done as described⁶. The *NRI* antisense riboprobe was labelled with [α -³²P]UTP. The hybridized sections were exposed to films, dipped in Kodak NTB3 nuclear emulsion, and after development counterstained with cresyl violet. Expression patterns and levels were ascertained by two sets of reactions using eight CxNR1KO and four *flox*/– control mice at P7 and two CxNR1KO and one *flox*/– control mice at P0. Sagittal sections from homozygous *NRI* null mutant were used as negative controls.

Histology

Whole embryos were fixed in 3.7% formaldehyde in 0.05 M phosphate buffer (pH 7.4) and stained overnight in LacZ solution (0.05 M phosphate buffer, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆, 2 mM MgCl₂, 1 mg ml^{–1} X-gal) at 37 °C. Brains (P0–8) were fixed, sectioned (400–500 μm) and stained overnight. In another set of experiments, P7 brains were sectioned (4–16 μm), mounted on slides, stained with X-gal, cresyl violet and/or GABA antibody (Sigma: A2052; diluted 1:5,000). Most experiments were done with positive (CAG-Δ-Z) and negative controls (CAG-CAT-Z) to monitor the reliability of the X-gal staining. The CAG-Δ-Z line was established by crossing CAG-CAT-Z reporter mice¹⁷ with deleter mice (CMV-Cre mice; a gift from A. Nagy) to remove the *loxP-CAT-loxP* fragment in the germline (not shown).

Cytochrome oxidase histochemistry, DiI labelling and cresyl violet staining procedures have been described⁶. For 5-HTT immunohistochemistry we used a rabbit polyclonal antibody (Diasorin; 1:10,000) and a secondary, biotinylated goat anti-rabbit antibody (Sigma, 1:200). For tenascin immunohistochemistry, we used a rabbit polyclonal primary antibody (a gift from K. Crossin, 1:500) and a CY-3 conjugated goat anti-rabbit secondary antibody (Chemicon, 1:100).

For areal quantification, 5-HTT immunostained flattened cortex sections were visualized under light microscope, and the images of the barrel field were acquired by CoolSnap digital camera (US Photometrics). Measurements of the entire large whisker-representation areas and individual 5-HTT immunopositive patches within that area were made using the Metaview Image Analysis Program (Universal Imaging). The ratio of total area of patches and the total whisker-representation area were determined for control and CxNR1KO cortex at P5.

Optical imaging

Cortical slices⁶ (400-μm thick) were kept in artificial cerebrospinal fluid (ACSF, in mM): 118 NaCl, 3 KCl, 2 CaCl₂, 1 MgCl₂, 1 NaH₂PO₄, 25 NaHCO₃, 10 glucose. Slices were stained with the voltage-sensitive dye, di-4-ANEPPS (5 μM, Molecular Probes) for 30 min and placed in an immersion-type recording chamber. An epifluorescence setup consisting of a ×1.6 objective, dichroic mirror (575 nm) and longpass filter (590 nm) was mounted above the slice. Fluorescence was excited with a laser (532 nm, Verdi, Coherent) and detected with a high speed CCD camera system (MiCAM, Brainwave) at a 1,333 Hz image rate. Optical responses represent averages of 16 stimulations (200 μs, 80–120 μA) at 0.05 Hz.

Received 27 April; accepted 5 July 2000.

- Woolsey, T. A. & Van der Loos, H. The structural organization of layer IV in the somatosensory region (SI) of the mouse cerebral cortex. *Brain Res.* **17**, 205–242 (1970).
- O'Leary, D. D. M., Ruff, N. L. & Dyck, R. H. Development, critical period plasticity, and adult reorganizations of mammalian somatosensory systems. *Curr. Biol.* **4**, 535–544 (1994).
- Chiaia, N. L., Fish, S. E., Bauer, W. R., Bennett-Clarke, C. A. & Rhoades, R. W. Postnatal blockade of cortical activity by tetrodotoxin does not disrupt the formation of vibrissa-related patterns in the rat's somatosensory cortex. *Dev. Brain Res.* **66**, 244–250 (1992).
- Schlaggar, B. L., Fox, K. & O'Leary, D. D. M. Postsynaptic control of plasticity in developing somatosensory cortex. *Nature* **364**, 623–626 (1993).
- Fox, K., Schlaggar, B. L., Glazewski, S. & O'Leary, D. D. M. Glutamate receptor blockade at cortical synapses disrupts development of thalamocortical and columnar organization in somatosensory cortex. *Proc. Natl Acad. Sci. USA* **93**, 5584–5589 (1996).
- Iwasato, T. *et al.* NMDA receptor-dependent refinement of somatotopic maps. *Neuron* **19**, 1201–1210 (1997).
- Woolsey, T. A. in *Development of Sensory Systems in Mammals*, (ed. Coleman, E. J.) 461–516 (Wiley, New York, 1990).
- Jhaveri, S. & Erzurumlu, R. S. in *Development of the Central Nervous System in Vertebrates* (eds Sharma, S. C. & Goffinet, A. M.) 167–178 (Plenum, New York, 1992).
- Constantine-Paton, M., Cline, H. T. & Debski, E. Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Annu. Rev. Neurosci.* **13**, 129–154 (1990).
- Goodman, C. S. & Shatz, C. J. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell* **72** (Suppl.), 77–98 (1993).
- Cramer, K. S. & Sur, M. Activity-dependent remodeling of connections in the mammalian visual system. *Curr. Opin. Neurobiol.* **5**, 106–111 (1995).
- Li, Y., Erzurumlu, R. S., Chen, C., Jhaveri, S. & Tonegawa, S. Whisker-related neuronal patterns fail to develop in the trigeminal brainstem nuclei of NMDAR1 knockout mice. *Cell* **76**, 427–437 (1994).
- Kutsuwada, T. *et al.* Impairment of sucking response, trigeminal neuronal pattern formation, and hippocampal LTD in NMDA receptor $\epsilon 2$ subunit mutant mice. *Neuron* **16**, 333–344 (1996).
- Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
- Guliano, M., Broccoli, V., Pardini, C. & Boncinelli, E. *Emx1* and *Emx2* show different patterns of expression during proliferation and differentiation of the developing cerebral cortex in the mouse. *Eur. J. Neurosci.* **8**, 1037–1050 (1996).
- Yoshida, M. *et al.* *Emx1* and *Emx2* functions in development of dorsal telencephalon. *Development* **124**, 101–111 (1997).
- Sakai, K. & Miyazaki, J.-I. A transgenic mouse line that retains Cre recombinase activity in mature oocytes irrespective of the cre transgene transmission. *Biochem. Biophys. Res. Commun.* **237**, 318–324 (1997).
- Erzurumlu, R. S. & Jhaveri, S. Thalamic axons confer a blueprint of the sensory periphery onto the developing rat somatosensory cortex. *Dev. Brain Res.* **56**, 229–234 (1990).
- Ottersen, O. P. & Storm-Mathisen, J. Glutamate- and GABA-containing neurons in the mouse and rat brain, as demonstrated with a new immunocytochemical technique. *J. Comp. Neurol.* **229**, 374–392 (1984).
- Parnavelas, J. G. The origin and migration of cortical neurons: new vistas. *Trends Neurosci.* **23**, 126–131 (2000).
- Tanifuji, M., Yamanaka, A., Sunaba, R., Terakawa, S. & Toyama, K. Optical responses evoked by white matter stimulation in rat visual cortical slices and their relation to neural activities. *Brain Res.* **738**, 83–95 (1996).

22. Feldmeyer, D., Egger, V., Lubke, J. & Sakmann, B. Reliable synaptic connections between pairs of excitatory layer 4 neurons within a single 'barrel' of developing rat somatosensory cortex. *J. Physiol. (Lond.)* **521**, 169–190 (1999).
23. Wong-Riley, M. T. & Welt, C. Histochemical changes in cytochrome oxidase of cortical barrels after vibrissal removal in neonatal and adult mice. *Proc. Natl Acad. Sci. USA* **77**, 2333–2337 (1980).
24. Cases, O. et al. Lack of barrels in the somatosensory cortex of monoamine oxidase A-deficient mice: Role of a serotonin excess during the critical period. *Neuron* **16**, 297–307 (1996).
25. Jhaveri, S., Erzurumlu, R. S. & Crossin, K. Barrel construction in rodent neocortex: Role of thalamic afferents versus extracellular matrix molecules. *Proc. Natl Acad. Sci. USA* **88**, 4489–4493 (1991).
26. Messersmith, E. K., Feller, M. B., Zhang, H. & Shatz, C. J. Migration of neocortical neurons in the absence of functional NMDA receptors. *Mol. Cell Neurosci.* **9**, 347–357. (1997).
27. Harris, R. M. & Woolsey, T. A. Dendritic plasticity in mouse barrel cortex following postnatal vibrissa follicle damage. *J. Comp. Neurol.* **196**, 357–376 (1981).
28. Katz, L. C. & Constantine-Paton, M. Relationship between segregated afferents and postsynaptic neurons in the optic tectum of three-eyed frogs. *J. Neurosci.* **8**, 3160–3180 (1988).
29. Rajan, I. & Cline, H. T. Glutamate receptor activity is required for normal development of tectal cell dendrites in vivo. *J. Neurosci.* **18**, 7836–7846 (1998).
30. Molnar, Z. & Hannan, A. J. in *Mouse Brain Development* (eds Goffinet, A. M. & Rakic, P.) 293–332 (Springer, Berlin, 2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Gerber for critical reading; M. Yoshida and S. Aizawa for information and a probe of *Emx1* gene; J. Tsien, Y. Li, J.-i. Miyazaki and A. Nagy for *NR1^{flax/flax}*, *NR1^{+/−}*, *CAG-CAT-Z* and *CMV-Cre* mice, respectively; C. Lovett for *Cre/LacZ* PCR primers; K. Crossin for tenascin antibody; M. Tanaka, H. Gomi, A. Kato, K. Ishii, T. Ikeda and H. Kanki for advice in experiments/manuscript preparation; T. A. Woolsey for helpful discussion; and N. Yoshida, Y. Onodera, R. Ando and R. Nomura for technical assistance. This work was supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan (T.L.), by the Whitehall Foundation and the NIH/NINDS (R.S.E.) and by the NIH (S.T.).

Correspondence and requests for materials should be addressed to S.I. (e-mail: sitohara@brain.riken.go.jp).

Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells

Zhen-Ming Pei^{*§}, Yoshiyuki Murata^{*§}, Gregor Benning[‡], Sébastien Thomine^{*}, Birgit Klüsener^{*}, Gethyn J. Allen^{*}, Erwin Grill[‡] & Julian I. Schroeder^{*}

^{*} Division of Biology, Cell and Developmental Biology, and Center for Molecular Genetics, University of California at San Diego, La Jolla, California 92093-0116, USA

[‡] Technische Universität München, Lehrstuhl für Botanik, D-85350 Freising-Weihenstephan, Germany

[§] These authors contributed equally to this work

Drought is a major threat to agricultural production. Plants synthesize the hormone abscisic acid (ABA) in response to drought, triggering a signalling cascade in guard cells that results in stomatal closure, thus reducing water loss¹. ABA triggers an increase in cytosolic calcium in guard cells ($[Ca^{2+}]_{cyt}$)^{2–6} that has been proposed to include Ca^{2+} influx across the plasma membrane^{3,5,7–9}. However, direct recordings of Ca^{2+} currents have been limited³ and the upstream activation mechanisms of plasma membrane Ca^{2+} channels remain unknown. Here we report activation of Ca^{2+} -permeable channels in the plasma membrane of *Arabidopsis* guard cells by hydrogen peroxide. The H_2O_2 -activated Ca^{2+} channels mediate both influx of Ca^{2+} in protoplasts and increases in $[Ca^{2+}]_{cyt}$ in intact guard cells. ABA induces the production of H_2O_2 in guard cells. If H_2O_2 production is blocked, ABA-induced closure of stomata is inhibited. Moreover, activation of Ca^{2+} channels by H_2O_2 and ABA- and H_2O_2 -induced stomatal closing are disrupted in the recessive ABA-

insensitive mutant *gca2*. These data indicate that ABA-induced H_2O_2 production and the H_2O_2 -activated Ca^{2+} channels are important mechanisms for ABA-induced stomatal closing.

Plasma membrane hyperpolarization in guard cells enhances ABA-induced increases in $[Ca^{2+}]_{cyt}$, indicating a possible hyperpolarization-activated Ca^{2+} influx component^{5,9}. In whole-cell patch-clamp recordings using conditions that enable Ca^{2+} current recordings in other plant cells^{10–12}, hyperpolarization did not activate Ca^{2+} currents in *Arabidopsis* guard cells ($n = 28$). Unexpectedly, when guard cells were exposed to H_2O_2 , membrane hyperpolarization consistently activated a large inward current ($n = 21$). To analyse the hyperpolarization-activated currents, we used Ba^{2+} ions, which are permeant to Ca^{2+} channels, with 0.1 mM dithiothreitol (DTT) in the pipette and bath solutions (see Methods). In the absence of H_2O_2 , only a small background conductance was found at membrane potentials from +32 to −198 mV (Fig. 1a, b; − H_2O_2). After addition of 5 mM H_2O_2 to the bath solution, a current was activated by hyperpolarization (Fig. 1a, b; $n = 22$). On average, currents at −198 mV were increased from -6.4 ± 0.3 pA to -72.9 ± 12.4 pA in response to 5 mM H_2O_2 ($P < 0.001$). Experiments using lower concentrations of H_2O_2 showed that the current amplitudes increase with increasing H_2O_2 concentrations (Fig. 1c). Currents at −198 mV were activated after addition of 50 μ M H_2O_2 ($P < 0.001$), indicating that physiological H_2O_2 concentrations¹³ could activate the currents. With Ca^{2+} in the bath solution, H_2O_2 -activated inward currents were observed in conjunction with rapid (R-type) anion channel^{14,15} activity (Fig. 1d, top). Substitution of Ca^{2+} with tetraethylammonium (TEA^+) in the bath solution abolished H_2O_2 -activated currents, without altering R-type anion currents (Fig. 1d, bottom). These data show that membrane stability was unaffected by H_2O_2 , and indicate that H_2O_2 -activated currents were carried by Ca^{2+} (Fig. 1d, top). Hyperpolarization-activated currents were not observed in *Arabidopsis* hypocotyl cells under the same conditions ($n = 27$; not shown), further indicating

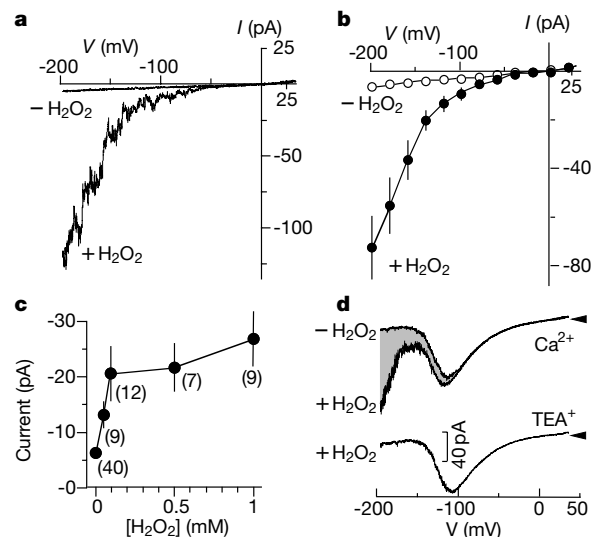


Figure 1 H_2O_2 activates inward currents in *Arabidopsis* guard cells. **a**, Whole-cell current recordings without (− H_2O_2) and with (+ H_2O_2) H_2O_2 . Voltage ramps were from +32 to −198 mV. Pipette and bath solutions contained 10 and 100 mM $BaCl_2$, respectively. **b**, Current–voltage relationship from cells as in **a** (+ H_2O_2 , $n = 40$ cells; − H_2O_2 , $n = 22$ cells). **c**, Average whole-cell currents at −198 mV against H_2O_2 concentration. The number of cells averaged is shown at each data point. Other conditions as in **a**. **d**, Whole guard-cell currents induced by voltage ramps from +37 to −193 mV without H_2O_2 (− H_2O_2) and with 5 mM H_2O_2 (+ H_2O_2). Top, recorded with 50 mM $CaCl_2$ in the bath (Ca^{2+}). Bottom, $CaCl_2$ replaced with 100 mM $TEA-Cl$ (TEA^+ ; $n = 18$). Shaded area, H_2O_2 -activated currents. Pipette solution contained 150 mM K_2SO_4 (ref. 15). Arrows on right show zero current levels.

that currents were not caused by non-specific H_2O_2 effects.

Over longer time periods, the H_2O_2 -activated current showed fluctuations between high and low activity states (Fig. 2a). We used changes in activity state to investigate the ionic selectivity of H_2O_2 -activated currents by applying rapid voltage ramps with a tenfold bath–cytosol BaCl_2 gradient (Fig. 2b). The reversal potential of $+17.1 \pm 2.3$ mV ($n = 14$) was close to the equilibrium potential for Ba^{2+} ($E_{\text{Ba}^{2+}} = +21.8$ mV, corrected for ionic activity), and far from the Cl^- equilibrium potential ($E_{\text{Cl}^-} = -54.0$ mV), showing that the H_2O_2 -activated currents are predominantly carried by Ba^{2+} . To test whether H_2O_2 -activated currents can be carried by different divalent cations, we replaced Ba^{2+} in the bath with Ca^{2+} , Mg^{2+} or TEA^+ . The H_2O_2 -activated currents showed large average Ca^{2+} , Ba^{2+} and Mg^{2+} currents, whereas TEA^+ was largely impermeant (Fig. 2c). Thus, H_2O_2 activates a Ca^{2+} -permeable, non-selective cation current (I_{Ca}) in *Arabidopsis* guard cells. At low I_{Ca} activity we observed discrete current steps, reminiscent of single-channel gating, which reversed at $+16.4 \pm 3.4$ mV ($n = 5$) under the same conditions as in Fig. 2b. We tested the effect of Ca^{2+} channel blockers on I_{Ca} (Fig. 2d). La^{3+} was an efficient and reversible inhibitor for I_{Ca} , whereas verapamil (1 μM) was less effective.

To determine whether I_{Ca} mediates cytosolic Ca^{2+} increases in guard cells, we simultaneously recorded I_{Ca} and $[\text{Ca}^{2+}]_{\text{cyt}}$ by combining whole-cell voltage-clamp recordings and ratiometric $[\text{Ca}^{2+}]_{\text{cyt}}$ measurements using the Ca^{2+} -sensitive dye fura-2 (ref. 3). Addition of H_2O_2 to the bath activated an inward current and increased $[\text{Ca}^{2+}]_{\text{cyt}}$ (Fig. 3a; $n = 8$). To determine the voltage dependencies of the H_2O_2 -induced current and $[\text{Ca}^{2+}]_{\text{cyt}}$ increases, we applied voltage ramps from $+34$ mV to -166 mV (Fig. 3b, c). Before H_2O_2 treatments, neither Ca^{2+} currents nor $[\text{Ca}^{2+}]_{\text{cyt}}$ changes were observed in response to hyperpolarization (Fig. 3b). After H_2O_2 treatment, hyperpolarizing voltage ramps activated currents

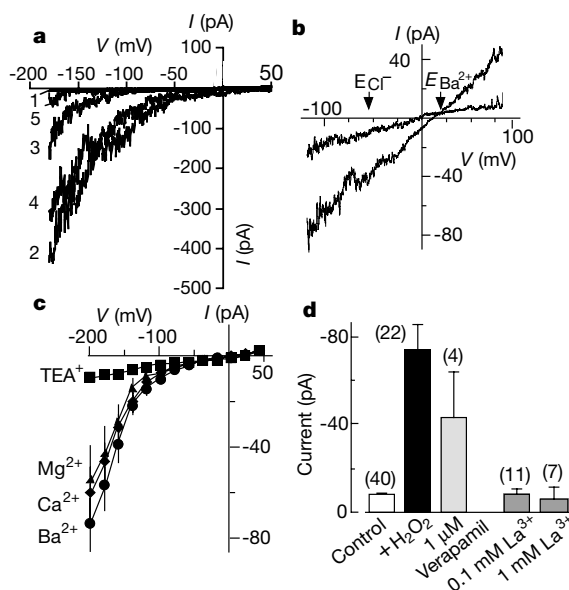


Figure 2 H_2O_2 -activated currents are Ca^{2+} -permeable, non-selective cation currents. **a**, Fluctuation in H_2O_2 -activated currents. 1 is 1 min before H_2O_2 addition; 2, 3, 4, 5 are 18, 24, 31 and 49 min after H_2O_2 application, respectively. Conditions as in Fig. 1a. **b**, Superimposed whole-cell currents recorded during rapid voltage ramps (0.7 mV s^{-1} ; -118 to $+82$ mV). Ramps were applied during active or inactive states of the hyperpolarization-activated current ($n = 14$ cells). Holding potential before voltage ramps, -168 mV. E_{Cl^-} , $E_{\text{Ba}^{2+}}$: Cl^- and Ba^{2+} equilibrium potentials, respectively. Standard pipette and bath solutions were used with 5 mM H_2O_2 in the bath. **c**, Current–voltage analyses recorded with 100 mM BaCl_2 ($n = 22$); 100 mM CaCl_2 ($n = 14$); 100 mM MgCl_2 ($n = 4$) or 200 mM TEA-Cl ($n = 12$) in the bath. 5 mM H_2O_2 was added to all bath solutions. **d**, Effect of Ca^{2+} channel blockers on H_2O_2 -activated Ca^{2+} currents at -180 mV. Indicated blocker concentrations and 5 mM H_2O_2 were added to the bath.

with a voltage dependence characteristic of I_{Ca} and concomitant $[\text{Ca}^{2+}]_{\text{cyt}}$ increases were recorded (Fig. 3c; $n = 3$). In addition, using a bath solution with competing K^+ ions containing 1 mM CaCl_2 and 20 mM KCl , I_{Ca} could still mediate $[\text{Ca}^{2+}]_{\text{cyt}}$ elevations ($n = 3$; not shown). To determine whether H_2O_2 affects $[\text{Ca}^{2+}]_{\text{cyt}}$ in intact *Arabidopsis* guard cells, we measured ratiometric changes of the GFP-based calcium reporter cameleon¹⁶. In epidermal peels bathed in 50 μM CaCl_2 and 5 mM KCl buffer, $[\text{Ca}^{2+}]_{\text{cyt}}$ increases in guard cells were consistently observed after addition of 100 μM H_2O_2 (Fig. 3d; $n = 16$ cells), as previously reported for H_2O_2 -induced $[\text{Ca}^{2+}]_{\text{cyt}}$ elevations in *Commelina*¹⁷. The average delay between H_2O_2 application and $[\text{Ca}^{2+}]_{\text{cyt}}$ elevation was 2.4 ± 0.8 min ($n = 8$), and the duration of $[\text{Ca}^{2+}]_{\text{cyt}}$ transients was 2.8 ± 0.5 min ($n = 8$). H_2O_2 application activated I_{Ca} within 2.7 ± 0.4 min ($n = 10$) with maximum delays of up to ~ 7 min. These delays are within the average delays for ABA-induced $[\text{Ca}^{2+}]_{\text{cyt}}$ elevations in hyperpolarized *Arabidopsis* guard cells of 12.6 ± 2.1 min (ref. 16) and up to 15 min in *Commelina*⁹. In addition, a delay in changing the oxidative state of guard cells upon H_2O_2 application can contribute to the observed lag time. In nominally zero extracellular Ca^{2+} (0.25 mM EGTA), no $[\text{Ca}^{2+}]_{\text{cyt}}$ increases were induced by extracellular H_2O_2 application ($n = 8$), further indicating that H_2O_2 triggers plasma membrane Ca^{2+} influx *in vivo*.

As $[\text{Ca}^{2+}]_{\text{cyt}}$ increases are an important signal in ABA-induced stomatal closure^{1,2}, we analysed whether H_2O_2 -activated Ca^{2+} currents contribute to guard cell ABA signalling. H_2O_2 at 100 μM triggered partial stomatal closing in *Arabidopsis* (Fig. 4a), similar to *Commelina*¹⁷. H_2O_2 at 10 μM also induced stomatal closing (not

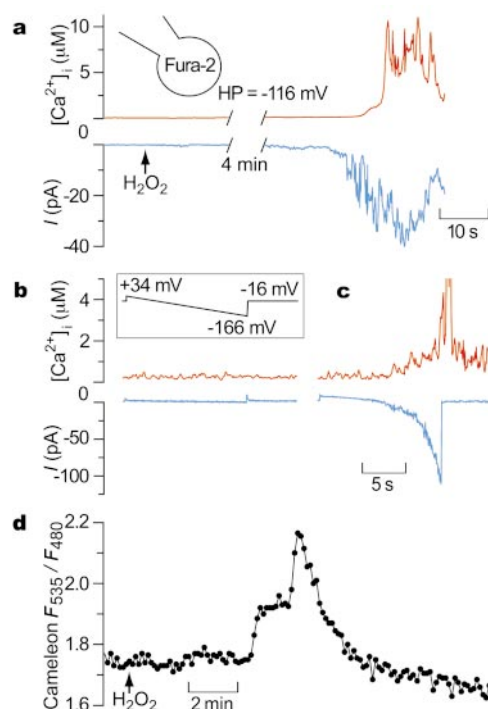


Figure 3 H_2O_2 -activated Ca^{2+} currents mediate Ca^{2+} influx in guard cell protoplasts and H_2O_2 -induced $[\text{Ca}^{2+}]_{\text{cyt}}$ increases in intact guard cells. **a**, Fura-2 ratios ($[\text{Ca}^{2+}]_{\text{cyt}}$) (red) and whole-cell currents (blue) recorded simultaneously in a guard cell. Holding potential, -116 mV. At the indicated time, 5 mM H_2O_2 was added. Pipette solution contained 0.5 mM fura-2 and 20 mM KCl ; bath solution contained 20 mM CaCl_2 . **b**, **c**, $[\text{Ca}^{2+}]_{\text{cyt}}$ elevations and whole-cell currents recorded simultaneously in a guard cell in response to voltage ramps (inset) before addition of H_2O_2 (**b**). The holding potential was -16 mV before voltage ramps allowing $[\text{Ca}^{2+}]_{\text{cyt}}$ to return to resting levels (**c**). $[\text{Ca}^{2+}]_{\text{cyt}}$ and currents recorded in response to a voltage ramp after H_2O_2 application. Representative traces in **a–c** were recorded in the same cell. **d**, Cameleon ratios ($[\text{Ca}^{2+}]_{\text{cyt}}$) measured in intact guard cells. Trace shows F_{535}/F_{480} ratio from one guard cell in response to H_2O_2 .

shown; $n = 40$ stomata; $P < 0.001$). Interestingly, H_2O_2 -induced stomatal closing required extracellular Ca^{2+} (Fig. 4a). To assess whether H_2O_2 is involved in ABA signalling, we blocked the production of reactive oxygen species (ROS) using diphenylene iodonium chloride (DPI), an inhibitor of NADPH oxidases^{13,18,19}. Figure 4b shows that 12.5 μM DPI partially inhibited ABA-induced stomatal closing. These data, together with the induction of I_{Ca} and $[\text{Ca}^{2+}]_{\text{cyt}}$ increases by H_2O_2 , and La^{3+} block of both I_{Ca} (Fig. 2c) and ABA-induced stomatal closing^{8,20}, led us to test whether H_2O_2 might be a second messenger in ABA signalling.

We monitored H_2O_2 generation in intact guard cells in epidermal strips treated with 50 μM ABA using a fluorescent dye, $\text{H}_2\text{DCF-DA}$, which reports changes in ROS in guard cells²¹. ABA treatment induced increases in fluorescence in guard cells (Fig. 4c), showing enhanced ROS accumulation. The relative fluorescence emission increased from 35.13 ± 2.15 ($n = 133$) to 52.36 ± 2.39 ($n = 162$) after treatment with 50 μM ABA (49.0% increase; $P < 0.001$). ABA-induced increases in ROS were observed in all experiments, including four blind experiments. Low concentrations of ABA (1 μM) also induced ROS production (36.8% increase; $n = 110$; $P < 0.001$). The initial onset of ABA-induced ROS production was further analysed. Figure 4d shows that ABA caused ROS elevation within ~ 1 min. Furthermore, ABA (50 μM) activated a hyperpolarization-induced Ba^{2+} current (Fig. 4e; $n = 13$), suggesting that I_{Ca} can contribute to ABA-induced $[\text{Ca}^{2+}]$ elevations and stomatal closing. Together, these data suggest a new mechanism within the ABA signal transduction cascade of guard cells, in which ABA induces H_2O_2 production, and H_2O_2 activates I_{Ca} , leading to Ca^{2+} influx and an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$. I_{Ca} and ABA-stimulated intracellular Ca^{2+} release mechanisms^{9,22,23} might mediate ABA-induced stomatal closing.

To test the hypothesis that H_2O_2 production and I_{Ca} are crucial to ABA signal transduction, we analysed *Arabidopsis* ABA signalling mutants. The recessive *gca2* mutant showed insensitivity of stomatal closing to ABA (G.B., T. Ehrler, M. Iten and E.G., unpublished data; and Fig. 5a)²⁴. Activation of I_{Ca} by H_2O_2 was disrupted in *gca2*

mutants (Fig. 5b, c; $n = 17$) as compared with wild-type controls (Fig. 2d). Concentrations of H_2O_2 up to 50 mM could not activate I_{Ca} in *gca2* guard cells (Fig. 5d; $n = 21$), in contrast to the wild type (Fig. 1d, top). These high concentrations of H_2O_2 did not interfere with R-type anion currents or membrane integrity in guard cells (Fig. 5d). Stomatal aperture measurements show that H_2O_2 -induced stomatal closing was also abolished in *gca2* mutants (Fig. 5e). ROS measurements showed that ABA at 50 μM increased the relative fluorescence emission in *gca2* guard cells ($43.6 \pm 8.2\%$ increase; $n = 50$; $P < 0.001$). Together, *gca2* analyses provide genetic evidence for the importance of H_2O_2 -activated I_{Ca} in ABA signalling. Furthermore, the findings that the *gca2* mutant impairs H_2O_2 activation of I_{Ca} (Fig. 5b–d) and H_2O_2 - and ABA-induced stomatal closing (Fig. 5a, e) indicate that *gca2* may impair signalling downstream of H_2O_2 production and upstream of or at the level of I_{Ca} (Fig. 5f).

We have shown that ABA causes ROS production and that ROS induce the activation of Ca^{2+} -permeable channels in the plasma membrane of *Arabidopsis* guard cells. We have also shown that these Ca^{2+} currents display a voltage-dependence with increasing activity at hyperpolarized potentials, as proposed⁵. Genetic, electrophysiological and physiological analyses with ROS measurements lead us to propose that ROS production and I_{Ca} function in the ABA signalling pathway in *Arabidopsis* guard cells (Fig. 5f). Furthermore, the findings that the recessive ABA-insensitive mutant *gca2* shows no H_2O_2 activation of I_{Ca} even at very high H_2O_2 concentrations (Fig. 5b–d) provides genetic evidence for the importance of I_{Ca} in ABA signal transduction (Fig. 5f). Identification of the *GCA2* gene will reveal the molecular basis of the defect in ABA signal transfer.

The properties of I_{Ca} differ from several types of Ca^{2+} -permeable channels characterized in plant cells^{10,12,25}. Characteristics of I_{Ca} correlate with recent reports of ABA- and hyperpolarization-evoked $[\text{Ca}^{2+}]_{\text{cyt}}$ increases in guard cells^{5,9,16}. I_{Ca} might share some similarities with hyperpolarization-activated Ca^{2+} -permeable channels in *Arabidopsis* root cells²⁶ and those activated by pathogen

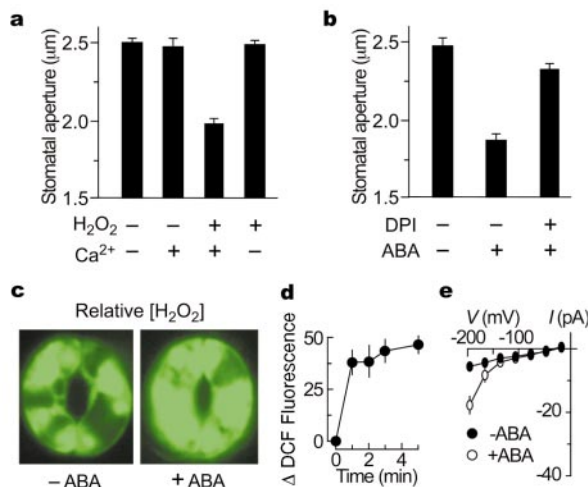


Figure 4 ABA causes H_2O_2 production and H_2O_2 -induced stomatal closing requires external Ca^{2+} . **a**, Stomatal closing measured with and/or without 0.1 mM extracellular free Ca^{2+} and/or 0.1 mM H_2O_2 . Data from 3 representative ($n = 60$ stomata per bar) of 12 experiments are shown. **b**, ABA-induced stomatal closure measured with and/or without 1 μM ABA and/or the NADPH oxidase inhibitor DPI (12.5 μM). **c**, ROS production in intact guard cells detected by ($\text{H}_2\text{DCF-AC}$) fluorescence in the absence or presence of 50 μM ABA. **d**, 25 μM ABA elicits ROS production within 1 min in guard cells. Average background fluorescence of DCF in control ethanol-treated guard cells was subtracted from each time point ($n = 5–21$ for controls and $n = 11–27$ for ABA-treated cells). **e**, ABA activation of I_{Ca} -type currents in *Arabidopsis* guard cells. 50 μM ABA was added to the bath ~ 10 min after beginning of whole-cell recordings, and ABA-activated currents were measured ~ 5 min after ABA exposure ($n = 13$). Data from the same cells are shown before (filled symbols) and after ABA application (open symbols).

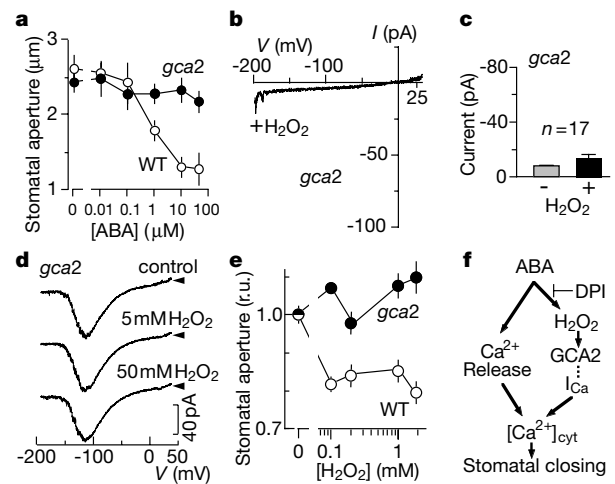


Figure 5 The recessive ABA-insensitive *Arabidopsis* mutant *gca2* disrupts H_2O_2 activation of Ca^{2+} currents and ABA-induced stomatal closing. **a**, ABA-induced stomatal closing in wild-type (WT) and *gca2* guard cells. Representative data from three ($n = 60$ stomata per data point) of six experiments are shown. **b**, Representative whole-cell current from a *gca2* guard cell treated with 5 mM H_2O_2 . Other conditions as in Fig. 1a. **c**, Effect of H_2O_2 on I_{Ca} in *gca2*. Data are mean of 17 cells. Compare with wild-type in Fig. 2d. **d**, Whole-cell depolarization-activated rapid anion currents recorded in a *gca2* guard cell under voltage ramps from +37 to -193 mV without (control, top) or with 5 mM and 50 mM H_2O_2 . Arrows indicate zero current levels. Other conditions as in Fig. 1d. **e**, Stomatal aperture closing as a function of H_2O_2 concentration ($n = 20–120$ stomata per data point). Apertures were normalized with respect to data in the absence of H_2O_2 (r.u., relative unit). **f**, Model for the role of H_2O_2 , I_{Ca} and *gca2* in the guard cell ABA signalling cascade. Additional ABA-induced intracellular Ca^{2+} release mechanisms are generalized in parallel for simplicity.

elicitors in tomato suspension cells¹¹, although the slow time dependence of the elicitor-activated currents differed from I_{Ca} . ABA (Fig. 4) and elicitors^{13,18,19} induce H_2O_2 production, and H_2O_2 has been shown to cause $[Ca^{2+}]_{cyt}$ elevations in plant-pathogen studies²⁷. Note that distinct elicitor-induced $[Ca^{2+}]_{cyt}$ elevation components have been implicated both before and after H_2O_2 production, indicating that more than one Ca^{2+} channel might contribute to this response^{12,13,27,28}. Pathogen elicitors cause H_2O_2 production in guard cells and stomatal closing²¹. Together these studies indicate that the H_2O_2 activation of the plasma membrane Ca^{2+} channel characterized here may represent central signalling components in plants, which integrate several stress signal transduction pathways that are known to cause $[Ca^{2+}]_{cyt}$ elevations²⁹. **Note added in proof:** Since submission of this work, a study has been published that shows hyperpolarization- and ABA-activated Ca^{2+} currents with characteristics similar to I_{Ca} in guard cells of *Vicia faba*³¹. □

Methods

Isolation of guard cells

Arabidopsis thaliana (Landsberg erecta and *gca2*) plants were grown in a controlled environment growth chamber with a 16/8-h light/dark cycle³⁰. We isolated guard cell protoplasts from 4- to 6-week-old plants as described³⁰.

Electrophysiology

Whole-cell voltage-clamp experiments on *Arabidopsis* guard cells were performed using Axopatch 1D, Axopatch 200 and 200A amplifiers (Axon Instruments) as described³⁰. Data were analysed using AXOGRAPH 3.5. Standard solutions for I_{Ca} measurements contained (in mM) 100 BaCl₂ (or 100 CaCl₂ or 100 MgCl₂), 0.1 DTT, 10 Mes-Tris, pH 5.6, in the bath and 10 BaCl₂ 0.1 DTT, 4 EGTA, 10 HEPES-Tris, pH 7.1, in the pipette. H_2O_2 (Sigma) was freshly added to bath solutions at the indicated concentrations. Without addition of 0.1 mM DTT to bath and pipette solutions, spontaneous I_{Ca} -like currents were occasionally recorded in guard cells. For ABA-activated Ca^{2+} current measurements (Fig. 4e), 1 mM NADPH was added to the standard pipette solution. The solutions used for rapid anion channel measurements contained (in mM) 150 K₂SO₄, 2 MgCl₂, 5 EGTA, 2.5 CaCl₂, 10 HEPES-Tris, pH 7.1, in the pipette, and 50 CaCl₂ or 100 TEA-Cl, 2 MgCl₂, 10 Mes-Tris, pH 5.6, in the bath. Other modifications to solutions are described in figure legends. Osmolalities of pipette and bath solutions were adjusted to 485 and 500 mmol kg⁻¹, respectively, using D-sorbitol.

Fura-2-based Ca^{2+} microphotometry

We recorded $[Ca^{2+}]_{cyt}$ and whole-cell currents simultaneously as described³. Guard cells were loaded with 0.5 mM fura-2 through patch clamp pipettes and ratiometric fluorescence measurements performed as described⁶. We recorded single-cell ratiometric signals and whole-cell currents simultaneously through an Axolab interface. $[Ca^{2+}]_{cyt}$ was calibrated *in vitro* using Ca^{2+} standards (0–39.8 μ M free Ca^{2+} ; Molecular Probes). The solutions used for fura-2 microphotometry contained (in mM) 0.5 fura-2 (pentapotassium salt; Molecular Probes), 20 KCl, 10 HEPES-Tris, pH 7.1, in the pipette, and 20 CaCl₂ (or 1 CaCl₂ and 20 KCl), 10 Mes-Tris, pH 5.6, in the bath.

Imaging of $[Ca^{2+}]_{cyt}$ in intact guard cells

We measured cytosolic Ca^{2+} using the GFP-based cameleon Ca^{2+} indicator in *Arabidopsis* guard cells as described¹⁶. Epidermal peels of rosette leaves, expressing a yellow cameleon construct, p35S-YC2.1-Kan (YC2.1), were placed in a microwell chamber containing 50 μ M CaCl₂ (or zero Ca^{2+} : 0 mM CaCl₂, 0.25 mM EGTA), 5 mM KCl and 10 mM Mes-Tris, pH 6.15. Excitation was provided at 440 nm, and emission ratiometric (F535 nm/F480 nm) images were collected using a Zeiss Axiovert microscope and METAFLUOR 2.75 software (Universal Imaging).

ROS detection in intact guard cells

We examined the production of ROS in guard cells by loading epidermal preparations with 2,7-dichlorofluorescein diacetate (H_2DCF -DA; Sigma) as described²¹. Epidermal strips were floated in 30 mM KCl, 10 mM Mes-KOH, pH 6.15, for 2 h. 50 μ M H_2DCF -DA was added to the floating solution for 10–30 min. For single time point measurements, we added 50 μ M ABA ($[+/-cis,trans$ -ABA; Sigma) 20 min before 10 min of dye loading. For time-resolved ROS measurements (Fig. 4d), we added ABA after dye loading. Control experiments, in which 0.1% ethanol was added, showed no stimulation of H_2O_2 production. Guard cells were observed under a fluorescent microscope (Eclipse E600, Nikon; excitation 450–490 nm, barrier 520–560 nm) equipped with a cooled CCD camera (Spots, Diagnostic Inc.). Images were acquired and the fluorescent emission of guard cells was analysed by using ADOBE PHOTOSHOP 5.0.

Stomatal aperture bioassays

Stomatal closing assays were conducted as described³⁰. Rosette leaves were floated in solutions containing (in mM) 10 KCl, 0.1 EGTA and 10 Mes-KOH, pH 6.15. Subsequently, 0.2 mM CaCl₂ (with 0.1 mM EGTA), 1 μ M ABA, 100 μ M H_2O_2 and 12.5 μ M diphenylene

iodonium chloride (DPI; Sigma) were added to the solutions as indicated. After treatment for 2 h, stomatal apertures were measured as described³⁰. All statistical analyses were performed using EXCEL 5.0 (Microsoft). Data are the mean $\pm$ s.e.m. Values of $P < 0.05$ were considered statistically significant.

Received 14 February; accepted 18 May 2000.

- MacRobbie, E. A. C. Signal transduction and ion channels in guard cells. *Phil. Trans. R. Soc. Lond. B* **353**, 1475–1488 (1998).
- McAinsh, M. R., Brownlee, C. & Hetherington, A. M. Absciscic acid-induced elevation of guard cell cytosolic Ca^{2+} precedes stomatal closure. *Nature* **343**, 186–188 (1990).
- Schroeder, J. I. & Hagiwara, S. Repetitive increases in cytosolic Ca^{2+} of guard cells by absciscic acid activation of nonselective Ca^{2+} permeable channels. *Proc. Natl Acad. Sci. USA* **87**, 9305–9309 (1990).
- Gilroy, S., Fricker, M. D., Read, N. D. & Trewavas, A. J. Role of calcium in signal transduction of *Commelina* guard cells. *Plant Cell* **3**, 333–344 (1991).
- Grabov, A. & Blatt, M. R. Membrane voltage initiates Ca^{2+} waves and potentiates Ca^{2+} increases with absciscic acid in stomatal guard cells. *Proc. Natl Acad. Sci. USA* **95**, 4778–4783 (1998).
- Allen, G. J., Kuchitsu, K., Chu, S. P., Murata, Y. & Schroeder, J. I. *Arabidopsis* *abi1-1* and *abi2-1* phosphatase mutations reduce absciscic acid-induced cytoplasmic calcium rises in guard cells. *Plant Cell* **11**, 1785–1798 (1999).
- Schwartz, A. Role of calcium and EGTA on stomatal movements in *Commelina communis*. *Plant Physiol.* **79**, 1003–1005 (1985).
- De Silva, D. L. R., Cox, R. C., Hetherington, A. M. & Mansfield, T. A. Synergism between calcium ions and absciscic acid in preventing stomatal opening. *New Phytol.* **101**, 555–563 (1985).
- Staxén, I. *et al.* Absciscic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. *Proc. Natl Acad. Sci. USA* **96**, 1779–1784 (1999).
- Piñeros, M. & Tester, M. Characterization of a voltage-dependent Ca^{2+} -selective channel from wheat roots. *Planta* **195**, 478–488 (1995).
- Gelli, A., Higgins, V. J. & Blumwald, E. Activation of plant plasma membrane Ca^{2+} -permeable channels by race-specific fungal elicitors. *Plant Physiol.* **113**, 269–279 (1997).
- Zimmermann, S. *et al.* Receptor-mediated activation of a plant Ca^{2+} -permeable ion channel involved in pathogen defense. *Proc. Natl Acad. Sci. USA* **94**, 2751–2755 (1997).
- Lamb, C. & Dixon, R. A. The oxidative burst in plant disease resistance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 251–275 (1997).
- Hedrich, R., Busch, H. & Raschke, K. Ca^{2+} and nucleotide dependent regulation of voltage dependent anion channels in the plasma membrane of guard cells. *EMBO J.* **9**, 3889–3892 (1990).
- Frachisse, J. M., Thomine, S., Colcombet, J., Guern, J. & Barbier-Brygoo, H. Sulfate is both a substrate and an activator of the voltage-dependent anion channel of *Arabidopsis* hypocotyl cells. *Plant Physiol.* **121**, 253–261 (1999).
- Allen, G. J. *et al.* Cameleon calcium indicator reports cytoplasmic calcium dynamics in *Arabidopsis* guard cells. *Plant J.* **19**, 735–747 (1999).
- McAinsh, M. R., Clayton, H., Mansfield, T. A. & Hetherington, A. M. Changes in stomatal behavior and guard cell cytosolic free calcium in response to oxidative stress. *Plant Physiol.* **111**, 1031–1042 (1996).
- Alvarez, M. E. *et al.* Reactive oxygen intermediates mediate a systemic signal network in the establishment of plant immunity. *Cell* **92**, 773–784 (1998).
- Orozco-Cardenas, M. & Ryan, C. A. Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. *Proc. Natl Acad. Sci. USA* **96**, 6553–6557 (1999).
- McAinsh, M. R., Brownlee, C. & Hetherington, A. M. Partial inhibition of ABA-induced stomatal closure by calcium-channel blockers. *Proc. R. Soc. Lond. B* **243**, 195–201 (1991).
- Lee, S. *et al.* Oligogalacturonic acid and chitosan reduce stomatal aperture by inducing the evolution of reactive oxygen species from guard cells of tomato and *Commelina communis*. *Plant Physiol.* **121**, 147–152 (1999).
- Leckie, C. P., McAinsh, M. R., Allen, G. J., Sanders, D. & Hetherington, A. M. Absciscic acid-induced stomatal closure mediated by cyclic ADP-ribose. *Proc. Natl Acad. Sci. USA* **95**, 15837–15842 (1998).
- Jacob, T., Ritchie, S., Assmann, S. M. & Gilroy, S. Absciscic acid signal transduction in guard cells is mediated by phospholipase D activity. *Proc. Natl Acad. Sci. USA* **96**, 12192–12197 (1999).
- Himmelbach, A., Iken, M. & Grill, E. Signalling of absciscic acid to regulate plant growth. *Phil. Trans. R. Soc. Lond. B* **353**, 1439–1444 (1998).
- Marshall, J., Corzo, A., Leigh, R. A. & Sanders, D. Membrane potential-dependent calcium transport in right-side-out plasma membrane vesicles from *Zea mays* L. roots. *Plant J.* **5**, 683–694 (1994).
- Kiegle, E., Gilliam, M., Haseloff, J. & Tester, M. Hyperpolarization-activated calcium currents found only in cells from the elongation zone of *Arabidopsis thaliana* roots. *Plant J.* **21**, 225–229 (2000).
- Price, A. H. *et al.* Oxidative signals in tobacco increase cytosolic calcium. *Plant Cell* **6**, 1301–1310 (1994).
- Knight, M. R., Campbell, A. K., Smith, S. M. & Trewavas, A. J. Transgenic plant aquaporin reports the effects of touch and cold-shock and elicitors on cytoplasmic calcium. *Nature* **352**, 524–526 (1991).
- Sanders, D., Brownlee, C. & Harper, J. F. Communicating with calcium. *Plant Cell* **11**, 691–706 (1999).
- Pei, Z.-M., Kuchitsu, K., Ward, J. M., Schwarz, M. & Schroeder, J. I. Differential absciscic acid regulation of guard cell slow anion channels in *Arabidopsis* wild-type and *abi1* and *abi2* mutants. *Plant Cell* **9**, 409–423 (1997).
- Hamilton, D. W. A., Hills, A., Köhler, B. & Blatt, M. R. Ca^{2+} channels at the plasma membrane of stomatal guard cells are activated by hyperpolarization and absciscic acid. *Proc. Natl Acad. Sci. USA* **97**, 4967–4972 (2000).

Acknowledgements

We thank J. Satimanont and C. Kwak for assistance; R. Tsien and M. Yanofsky for use of imaging systems and fluorescence microscopes; and J. Kwak, D. Lee, R. Vaidyanathan, M. Ghassemin, N. Leonhardt and P. Buschmann for comments on the manuscript. This work was supported by Department of Energy, National Institutes of Health and National Science Foundation grants to J.I.S. and by a European Biotech Program grant and Fonds der Chemischen Industrie to E.G. Y.M. was supported by a fellowship for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, S.T. and G.A. by Human Frontier Science Program fellowships, and B.K. by a Feodor-Lynen fellowship from the Alexander von Humboldt Foundation.

Correspondence and requests for materials should be addressed to J.I.S. (e-mail: julian@biomail.ucsd.edu).

Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase

John D. McKinney^{*,†}, Kerstin Höner zu Bentrup^{‡,§}, Ernesto J. Muñoz-Elias[§], Andras Miczak^{†,||}, Bing Chen^{*}, Wai-Tsing Chan[§], Dana Swenson^{†,||}, James C. Sacchettini^{||}, William R. Jacobs, Jr^{*} & David G. Russell^{†,||}

^{*} Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, New York 10461, USA

[‡] Department of Molecular Microbiology, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, Missouri 63110, USA

[§] The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

^{||} Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, USA

[†] Present addresses: The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA (J.D.M. and W.-T.C.); Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA (K.H.z.B., D.S. and D.G.R.); Department of Microbiology, Albert Szent-Gyorgyi Medical University, Szeged POB 8-6701, Hungary (A.M.).

Mycobacterium tuberculosis claims more human lives each year than any other bacterial pathogen. Infection is maintained in spite of acquired immunity and resists eradication by antimicrobials^{1,2}. Despite an urgent need for new therapies targeting persistent bacteria, our knowledge of bacterial metabolism throughout the course of infection remains rudimentary. Here we report that persistence of *M. tuberculosis* in mice is facilitated by isocitrate lyase (ICL), an enzyme essential for the metabolism of fatty acids^{3,4}. Disruption of the *icl* gene attenuated bacterial persistence and virulence in immune-competent mice without affecting bacterial growth during the acute phase of infection. A link between the requirement for ICL and the immune status of the

host was established by the restored virulence of Δicl bacteria in interferon- γ knockout mice. This link was apparent at the level of the infected macrophage: Activation of infected macrophages increased expression of ICL, and the Δicl mutant was markedly attenuated for survival in activated but not resting macrophages. These data suggest that the metabolism of *M. tuberculosis* *in vivo* is profoundly influenced by the host response to infection, an observation with important implications for the treatment of chronic tuberculosis.

The acquisition of essential nutrients by intracellular pathogens remains an area of considerable interest and ignorance. Biochemical studies suggest that in chronically infected lung tissues, fatty acids might be a major source of carbon and energy in *M. tuberculosis* metabolism⁵. Two pathways are required for fatty acid use by bacteria: the β -oxidation cycle and the glyoxylate shunt^{3,4}. The glyoxylate shunt is essential for carbon anaplerosis in the Krebs cycle during growth on C₂ substrates such as fatty acids, which are the only abundant C₂ carbon sources in mammalian tissues⁶. The glyoxylate shunt is widespread among prokaryotes lower eukaryotes and plants, but is apparently absent in vertebrates.

Recent studies revealed that expression of isocitrate lyase (ICL), an enzyme of the glyoxylate shunt, is upregulated during infection of macrophages by *Mycobacterium* spp.^{7–9}. Genes encoding isocitrate lyase activity in mycobacteria were identified by a genetic approach. A mutant of *M. smegmatis* was isolated which failed to grow on C₂

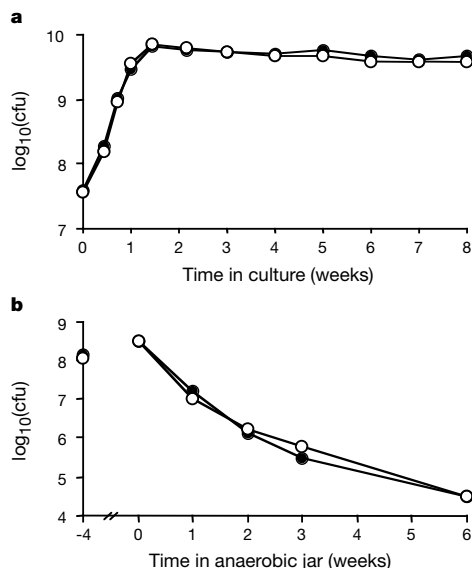


Figure 1 Absence of ICL does not affect growth or stationary phase survival *in vitro*. **a**, **b**, Wild-type (circles) and Δicl mutant (filled circles) bacteria were enumerated by plating for colony-forming units (cfu). Samples at each time point were assayed in triplicate and averaged; standard deviations were less than 10% of the mean. **a**, Log-phase growth and stationary-phase survival. Bacteria were grown in liquid medium with aeration at 37 °C for 8 weeks. **b**, Microaerophilic growth and anaerobic survival. Bacteria were grown in liquid medium without aeration at 37 °C for 4 weeks, then transferred to anaerobic jars and incubated at 37 °C for 6 weeks.

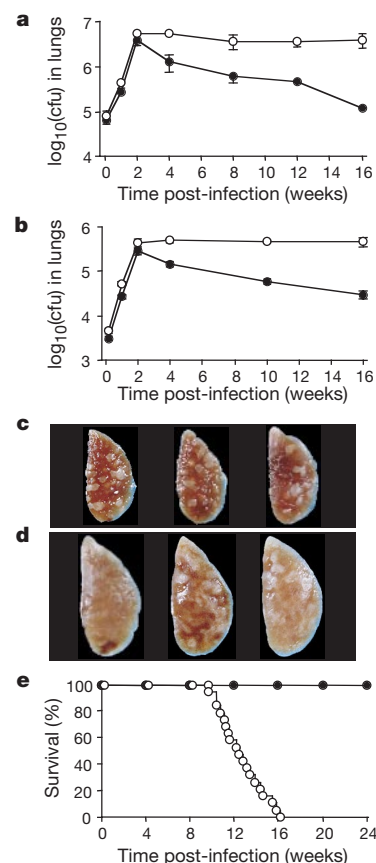


Figure 2 ICL is required for persistence and virulence of *M. tuberculosis* in immune-competent mice. **a**, **b**, Bacteria in the lungs of intravenously infected mice were enumerated by plating for cfu. Error bars indicate standard errors for each group ($n = 4$ or 5). **a**, Bacterial loads in B6X129-F₁ mice infected with 10⁶ cfu of Δicl (filled circles) or wild-type (circles) bacteria. **b**, Bacterial loads in C57BL/6 mice infected with 10⁵ cfu of Δicl bacteria transformed with pMV261 (filled circles) or pMV261 expressing ICL (circles). **c**, **d**, Lung pathology in B6X129-F₁ mice infected with Δicl (**c**) or wild-type (**d**) bacteria at 16 weeks. **e**, Survival of Balb/c mice ($n = 19$ per group) infected with 10⁶ cfu of Δicl (filled circles) or wild-type (circles) bacteria.

substrates unless rescued by the *Escherichia coli aceA* gene encoding isocitrate lyase. *M. tuberculosis* expresses two enzymes with isocitrate lyase activity, ICL (Rv0467) and AceA (Rv1915/6)⁷. However, only the gene encoding ICL and not AceA was able to rescue the *M. smegmatis icl* mutant for growth on C₂ substrates (data not shown). The *icl* gene was disrupted in the virulent Erdman strain of *M. tuberculosis* through allelic exchange and the Δicl mutation was confirmed by Southern blotting (data not shown).

In *E. coli* and *Salmonella typhimurium*, induction of the fatty acid catabolic regulon is essential for survival in stationary phase^{10,11}, and a similar role has been proposed in *M. tuberculosis*. In one study, ICL activity was induced when well-aerated cultures of *M. tuberculosis* grew to saturation¹², which suggests a role for ICL in stationary-phase survival. However, we observed no phenotype for the Δicl mutant of *M. tuberculosis* in log phase or stationary phase *in vitro* (Fig. 1a). In another study, *M. tuberculosis* ICL was induced by oxygen limitation¹³, and it was suggested that ICL might contribute to adaptation to hypoxia. However, we found that wild-type and Δicl strains of *M. tuberculosis* were indistinguishable phenotypically when cultured in hypoxic or anoxic atmospheres (Fig. 1b).

The contribution of ICL to *in vivo* metabolism of *M. tuberculosis* was assessed by infecting immune-competent mice with wild-type or Δicl bacteria (Fig. 2). The Δicl mutation had little effect on bacterial growth during the acute phase of infection (0–2 weeks): Mean doubling times were 50.8 h for the wild type versus 52.6 h for Δicl bacteria (Fig. 2a). However, from 2 weeks onwards, the Δicl mutant was eliminated progressively from the lungs (and extrapulmonary organs, data not shown), with a more than 1.5 log decline in bacterial burden by 16 weeks. In contrast, the peak bacterial load of wild-type *M. tuberculosis* was maintained for the duration of the experiment. There is some evidence that the plateau in bacterial numbers during persistent infection reflects bacterial stasis rather than balanced growth and killing^{14,15}, but this point remains controversial. The persistence defect caused by the Δicl mutation was not due to polar effects on neighbouring genes because the wild-type phenotype was restored by a plasmid expressing ICL (Fig. 2b).

The reduced ability of the Δicl mutant to sustain an infection was accompanied by attenuated virulence (Fig. 2c–e). At 2 weeks post-

infection, lungs of mice infected with wild-type and Δicl bacteria showed macroscopic lesions that were comparable in size and number (data not shown), reflecting the similarity in bacterial loads at 2 weeks (Fig. 2a). By 16 weeks, however, disease progression had diverged dramatically; the lungs of mice infected with Δicl bacteria showed little change between 2 and 16 weeks (Fig. 2c), whereas the lungs of mice infected with wild-type bacteria became grossly inflamed and enlarged, with numerous expanding and coalescing tubercles (Fig. 2d). Attenuated virulence of the Δicl mutant was also demonstrated by infection of Balb/c mice, which exhibit a more rapid disease progression. Balb/c mice infected with wild-type bacteria succumbed between days 68 to 113 (average, 88 days), whereas all mice infected with Δicl bacteria were surviving at day 168 (Fig. 2e).

These data show that ICL is important for survival of *M. tuberculosis* in the lungs of mice during the persistent phase of infection. In contrast, ICL is dispensable for bacterial growth in the acute phase of infection. These observations imply a 'C₂ shift' in bacterial carbon metabolism concomitant with the host's response to infection. The emergence of adaptive immunity *in vivo* is paralleled by the accumulation of inflammatory macrophages in the lungs. In resting macrophages, pathogenic mycobacteria replicate exponentially within vacuoles that are blocked in maturation, acidification, and fusion with lysosomes^{16–19}. In contrast, bacteria that enter activated macrophages confront a more hostile environment, in which vacuole maturation and acidification, restricted access to nutrients, and reduced oxygen tension may all constrain bacterial growth^{20–22}. These environmental changes could affect the metabolism of intracellular mycobacteria. Therefore, we examined whether the activation status of the infected macrophage influenced the expression of and requirement for ICL.

To analyse the effect of macrophage activation on ICL expression we constructed a bifunctional ICL–GFP fusion that was fluorescent yet retained ICL activity. The *icl-gfp* gene was expressed from the *icl* promoter on a plasmid or by replacement of the chromosomal *icl* gene by homologous recombination²³. Expression of ICL–GFP was induced by palmitate and repressed by succinate (data not shown), consistent with our results for the wild-type ICL enzyme⁷. The relative levels of ICL–GFP were evaluated by fluorescence microscopy

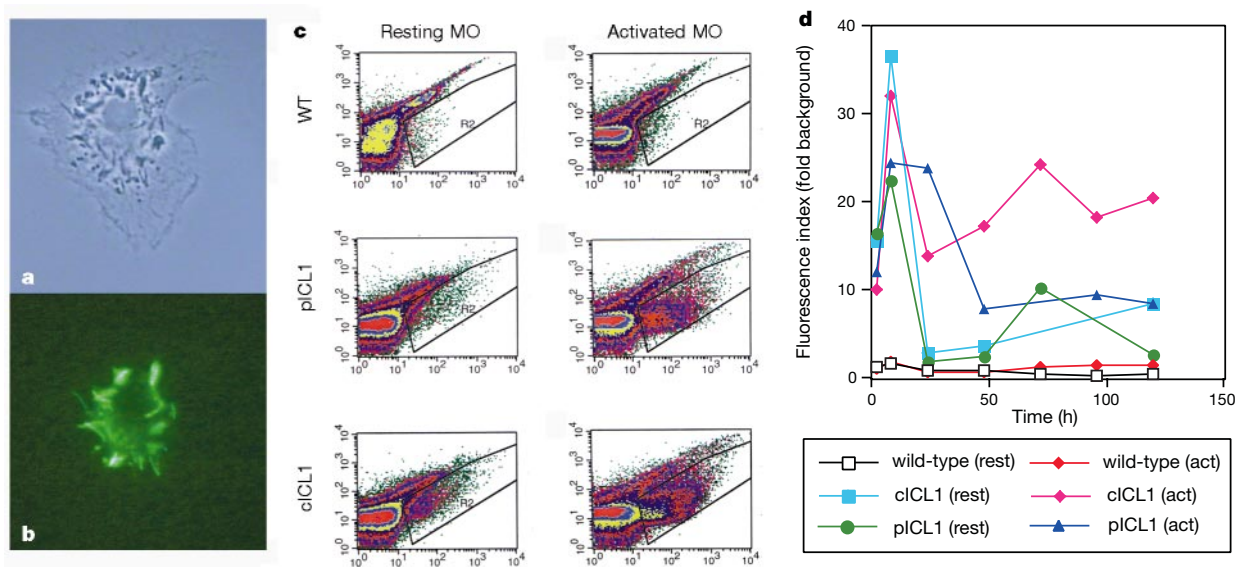


Figure 3 The level of ICL–GFP is higher in bacteria within activated macrophages. **a, b**, Light (**a**) and fluorescence (**b**) micrographs of *M. tuberculosis* (pICL–GFP) 24 h post-infection in macrophages pre-activated with IFN- γ and LPS. **c**, Flow cytometry of bacteria from resting and activated macrophages infected 48 h previously with wild-type *M. tuberculosis* (WT), bacteria expressing ICL–GFP from a plasmid (pICL1), or bacteria

with the chromosomal copy of *icl* replaced by the *icl-gfp* reporter (cICL1). Fluorescent indices were calculated from the gated areas in each panel. **d**, A graph illustrating the fluorescent indices of WT, pICL1 and cICL1 bacteria from resting and activated macrophages over 120 h. The x axis is relative fluorescence and the y axis is side scatter; both are log scales.

and flow cytometry following infection of macrophages that were non-activated or pre-activated with interferon- γ (IFN- γ) and lipopolysaccharide (LPS) (Fig. 3). In both resting and activated macrophages, we observed elevated levels of ICL-GFP shortly after infection. However, the levels of ICL-GFP in the bacilli returned to background after 24 h in resting macrophages, and yet remained elevated in activated macrophages for the duration of the experiment.

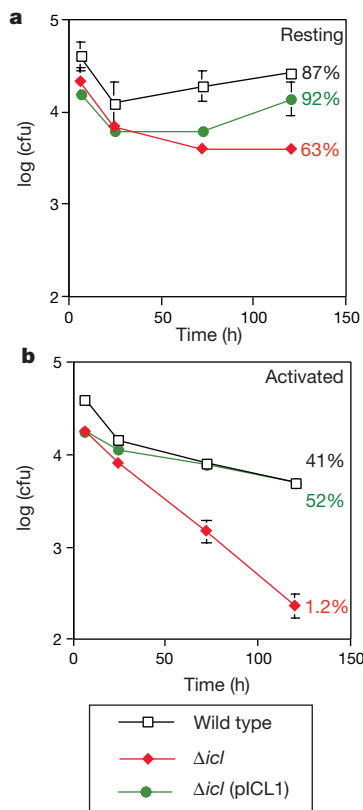


Figure 4 The requirement for ICL is determined by the immune status of the host cell. **a, b**, Bone marrow murine macrophages were not activated (**a**) or activated by 16-h exposure to 50 U ml⁻¹ rIFN- γ (**b**) before infection with wild-type *M. tuberculosis*, *M. tuberculosis* Δicl , or *M. tuberculosis* Δicl complemented with ICL-GFP (pICL1) at a multiplicity of 10:1 (bacteria per macrophage). The numbers in brackets following the final time point refer to the percentage survival of each strain relative to infection at the 6-h time point.

The observation that ICL expression is linked to the activation status of the infected host cell suggests that ICL might be more important for bacterial survival in activated than in resting macrophages. This idea paralleled our finding that ICL is required for bacterial survival specifically during the persistent phase of infection in mice. We therefore compared the phenotype of the *M. tuberculosis* Δicl mutant in resting versus activated macrophages at 6, 24, 72 and 120 h post-infection (Fig. 4). All preparations showed a marked decrease in bacterial numbers between 6 and 24 h irrespective of bacterial strain or host-cell status. However, after 24 h, the different bacterial phenotypes became apparent. In resting macrophages, the Δicl mutant showed a modest reduction in viability in comparison to the wild-type and pICL1-complemented Δicl strains (Fig. 4a). In activated macrophages, however, the survival of the Δicl mutant was markedly impaired (1.2%) relative to both the wild type (41%) and the complemented Δicl mutant (52%) (Fig. 4b). This trend was confirmed in three independent experiments.

Our observation that the phenotype of Δicl bacteria was more pronounced in activated than resting macrophages suggests a direct link between the immune status of the host and the requirement for ICL. Further evidence for this association comes from the restored virulence of the Δicl mutant on infection of immune-deficient, IFN- γ knockout mice (Fig. 5). The Δicl mutant grew progressively in the lungs of IFN- γ ^{-/-} mice (Fig. 5a) and infection was rapidly lethal (Fig. 5b). A slight residual phenotype was noted for the Δicl mutant in the IFN- γ ^{-/-} mice; bacterial growth slowed after 10 days of infection and lethality was somewhat delayed, which might be due to other macrophage-activating cytokines such as tumour necrosis factor- α . Our findings establish a link between the immune status of the host and the metabolism of *M. tuberculosis* *in vivo* that is mediated through activation of parasitized macrophages.

Persistence of bacteria and chronicity of infection are hallmarks of tuberculosis. Patients with chronic tuberculosis are thought to harbour bacteria in various metabolic states, ranging from active cell growth and division to stationary phase^{24,25}. Conventional drugs target processes required for bacterial cell growth and division, such as cell-wall biogenesis and chromosome replication^{1,2,25}. Poor activity against slow- or non-growing bacteria is thought to be an important reason why conventional drugs take so long to eradicate infection^{24,25}. Therefore, our demonstration that ICL promotes persistence of infection by enhancing bacterial survival within inflammatory macrophages makes it an attractive new target for chemotherapy. The development of ICL inhibitors will be facilitated

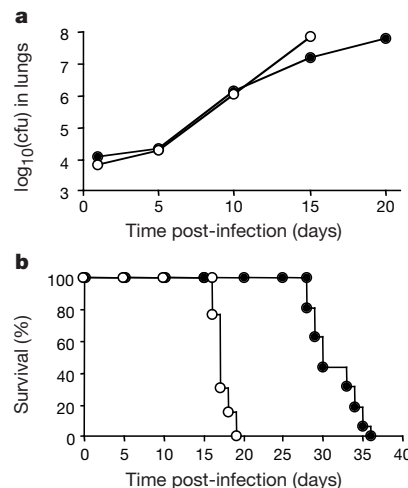


Figure 5 The Δicl mutant is lethal to immune-deficient mice. IFN- γ ^{-/-} mice were infected intravenously with 10⁶ cfu Δicl (filled circles) or wild-type (circles) bacteria. **a**, Lungs were collected at the indicated time points and bacterial loads were quantified by plating for cfu.

Error bars indicate standard errors for each group ($n = 5$). **b**, Survival of infected IFN- γ ^{-/-} mice was, on average, 17 days (wild-type bacteria) and 33 days (Δicl /bacteria), $n = 16$ per group.

by the recent solution of the three-dimensional structure of *M. tuberculosis* ICL in association with the prototypic inhibitors 3-bromopyruvate and 3-nitropropionate³¹. Future efforts will focus on the development of ICL inhibitors as new drug candidates with preferential activity against persistent bacteria. □

Methods

Mycobacterial strains and growth conditions

M. tuberculosis (Erdman and CSU93) were passaged once through mice and stored in aliquots at -80°C . *M. smegmatis* mc²155 (ref. 26) was colony-purified and stored in aliquots at -80°C . Mycobacteria were grown in 7H9 broth or 7H10 agar, with 10% OADC media supplement (DifCo), 0.5% glycerol, 100 $\mu\text{g ml}^{-1}$ cycloheximide, and 0.1% Tween-80. Antibiotics were hygromycin at 50 $\mu\text{g ml}^{-1}$ or kanamycin at 25 $\mu\text{g ml}^{-1}$. Defined carbon medium was M9 agar (DifCo) supplemented with glucose, sodium acetate, or methyl palmitate at 0.1%. Microaerophilic growth and anaerobic survival of *M. tuberculosis* were as described¹³.

Isolation and complementation of an *icl* mutant of *M. smegmatis*

M. smegmatis mc²155 was mutated with 2.5% ethyl methane sulphonate (Sigma) in 0.1 M phosphate buffer (pH 7) for 90 min, washed, recovered in 7H9 broth for 6 h at 37°C , and plated for colonies on 7H10 agar. Two *icl* mutants were identified as colonies that failed to grow on M9 agar + 0.1% sodium acetate unless transformed with pJM007, which contained a fusion of the *E. coli* *aceA* gene (gift of B. McFadden) to the mycobacterial heat shock promoter in pMV261²⁷. The *M. tuberculosis* *icl* gene was isolated from a mycobacterial genomic library (gift of F. C. Bange) by marker rescue.

Disruption of *icl* in *M. tuberculosis*

A 2.7-kbp (kilobase pairs) *Bam*HI-*Cl*AI genomic fragment spanning the *M. tuberculosis* *icl* gene was inserted into pYUB631²⁸ and a 685 bp *Xho*I fragment internal to *icl* was replaced with the *hyg* hygromycin resistance cassette. The deletion/disruption eliminated the codons for amino acids 65–290 of the 428-amino-acid ICL protein and abolished biological activity. The Δ icl:hyg allele was substituted for *icl* in the *M. tuberculosis* chromosome by allelic exchange, as described²⁸, and confirmed by Southern blot probed with a 981-bp *Sac*II fragment from the *M. tuberculosis* *icl* open reading frame (ORF).

M. tuberculosis growth and persistence in mice

C57BL/6J, 129SEvEv, Balb/c and IFN γ ^{−/−} mice were purchased from Jackson Laboratories. Frozen stocks of *M. tuberculosis* strains were thawed, diluted to about 10^7 or 10^6 colony-forming units (cfu) per ml in PBS/Tween, and sonicated. Mice were infected by tail-vein injection of 0.1 ml (about 10^6 or 10^5 cfu) of the bacterial suspension. Infected mice ($n = 4$ or 5 per group) were anesthetized by intraperitoneal injection of 0.1 ml of 10 mg ml^{−1} Nembutal and killed by dislocation of the cervical vertebrae. Organs were transferred to plastic Tekmar bags with 10 ml PBS/0.1% Tween-80 and homogenized in a Tekmar Stomacher. Organ homogenates were diluted and plated on 7H10 agar. Colonies were scored after 3–4 weeks at 37°C . Portions of tissues were fixed in phosphate-buffered formalin (10%) and photographed.

Reporter construction and analysis of ICL–GFP expression

pICL1 (*icl*–*gfp*) was derived from pMV262 (ref. 27) by substitution of the *icl* promoter (310-bp 5' of the start codon of *icl*) and the *icl* ORF fused at the 3' end to the *mut2* green fluorescent protein gene²⁹. The *icl*–*gfp* fusion gene was also used to replace the chromosomal *icl* gene through homologous recombination (cICL1) in *M. tuberculosis* CSU93 (ref. 23). Quantification of bacterial ICL–GFP expression in infected macrophages was obtained by flow cytometry. Murine bone-marrow macrophages were non-activated or activated with IFN- γ (100 U ml^{−1}, 16 h) and LPS (1 $\mu\text{g ml}^{-1}$, 2 h). Infected monolayers were washed three times with PBS and lysed with 0.1% Saponin. Intact cells and nuclei were pelleted (250 $\times$ g, 5 min) and the supernatants containing bacteria were collected. Bacteria were pelleted (2,000 $\times$ g, 20 min), fixed with 4% paraformaldehyde in PBS for 30 min, washed once with PBS/0.05% Tween-80/0.1% BSA, and re-suspended in PBS/Tween. Flow cytometry was done with a FACScalibur cytometer (Becton Dickinson Immunocytometry Systems) as described³⁰ and data were collected on 5×10^4 bacterium-sized particles per sample. Background fluorescence was determined by analysis of macrophages infected with wild-type bacteria. Relative fluorescence was expressed as fold induction over background.

Survival of Δ icl bacteria in resting and activated macrophages

Murine bone-marrow macrophages were either non-activated or activated with rIFN- γ (50 U ml^{−1}, 16 h). This activation protocol promotes acidification of bacteria-containing vacuoles without strong induction of inducible nitric oxide synthase (NOS2), a potent antimicrobial response. The *M. tuberculosis* Erdman strains (wild-type, Δ icl mutant, and Δ icl mutant complemented with pICL1) were added at a multiplicity of 10:1 (bacteria per macrophage) to duplicate wells, incubated for 2 h, washed, and fresh medium was added. Samples were collected at 6, 24, 72 and 120 h time points by lysing infected cells with 0.5% Tween-20. Lysates were diluted in PBS/Tween and plated on 7H10 agar. Colonies were scored after 3–4 weeks at 37°C .

Received 25 February; accepted 7 June 2000.

- McKinney, J. D., Jacobs, W. R. & Bloom, B. R. in *Emerging Infections* (eds Krause, R., Gallin, J. I. & Fauci, A. S.) 51–146 (Academic, New York, 1998).
- Parrish, N. M., Dick, J. D. & Bishai, W. R. Mechanisms of latency in *Mycobacterium tuberculosis*. *Trends Microbiol.* **6**, 107–112 (1998).
- Clark, D. P. & Cronan, J. E. Jr in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 343–357 (ASM Press, Washington DC, 1996).
- Cronan, J. E. Jr. & LaPorte, D. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 206–216 (ASM Press, Washington DC, 1996).
- Segal, W. in *The Mycobacteria: A Sourcebook* (eds Kubica, G. P. & Wayne, L. G.) 547–573 (Dekker, New York, 1984).
- Wheeler, P. R. & Ratledge, C. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 353–385 (ASM Press, Washington DC, 1994).
- Höner zu Bentrup, K., Miczak, A., Swenson, D. L. & Russell, D. G. Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J. Bacteriol.* **181**, 7161–7167 (1999).
- Sturgill-Koszycki, S., Haddix, P. L. & Russell, D. G. The interaction between *Mycobacterium* and the macrophage analyzed by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* **18**, 2558–2565 (1997).
- Graham, J. E. & Clark-Curtiss, J. E. Identification of *Mycobacterium tuberculosis* RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). *Proc. Natl Acad. Sci. USA* **96**, 11554–11559 (1999).
- Farewell, A., Diez, A. A., DiRusso, C. C. & Nystrom, T. Role of the *Escherichia coli* FadR regulator in stasis survival and growth phase-dependent expression of the *uspA*, *fad*, and *fab* genes. *J. Bacteriol.* **178**, 6443–6450 (1996).
- Spector, M. P. *et al.* The medium-/long-chain fatty acyl-CoA dehydrogenase (fadF) gene of *Salmonella typhimurium* is a phase 1 starvation-stress response (SSR) locus. *Microbiol.* **145**, 15–31 (1999).
- Suryanarayana Murthy, P., Sirsi, M. & Ramakrishnan, T. Effect of age on the enzymes of tricarboxylic acid and related cycles in *Mycobacterium tuberculosis* H37Rv. *Amer. Rev. Resp. Dis.* **108**, 689–690 (1973).
- Wayne, L. G. & Lin, K. Y. Glyoxylate metabolism and adaptation of *Mycobacterium tuberculosis* to survival under anaerobic conditions. *Infect. Immunol.* **37**, 1042–1049 (1982).
- Rees, R. J. W. & Hart, P. D. Analysis of the host-parasite equilibrium in chronic murine tuberculosis by total and viable bacillary counts. *Br. J. Exp. Pathol.* **42**, 83–88 (1961).
- Wallace, J. G. The heat resistance of tubercle bacilli in the lungs of infected mice. *Am. Rev. Respir. Dis.* **83**, 866–871 (1961).
- Hart, P. D. & Young, M. R. Interference with normal phagosome-lysosome fusion in macrophages, using ingested yeast cells and suramin. *Nature* **256**, 47–49 (1975).
- Sturgill-Koszycki, S. *et al.* Lack of acidification in *Mycobacterium* phagosomes produced by exclusion of the vesicular proton-ATPase. *Science* **263**, 678–681 (1994).
- Sturgill-Koszycki, S., Schaible, U. E. & Russell, D. G. *Mycobacterium*-containing phagosomes are accessible to early endosomes and reflect a transitional state in normal phagosome biogenesis. *EMBO J.* **15**, 6960–6968 (1996).
- Clemens, D. L. & Horwitz, M. A. The *Mycobacterium tuberculosis* phagosome interacts with early endosomes and is accessible to exogenously administered transferrin. *J. Exp. Med.* **184**, 1349–1355 (1996).
- James, P. E., Grinberg, O. Y., Michaels, G. & Swartz, H. M. Intraphagosomal oxygen in stimulated macrophages. *J. Cell. Physiol.* **163**, 241–247 (1995).
- Schaible, U. E., Sturgill-Koszycki, S., Schlesinger, P. H. & Russell, D. G. Cytokine activation leads to acidification and increases maturation of *Mycobacterium avium*-containing phagosomes in murine macrophages. *J. Immunol.* **160**, 1290–1296 (1998).
- Via, L. E. *et al.* Effects of cytokines on mycobacterial phagosome maturation. *J. Cell Sci.* **111**, 897–905 (1998).
- Pellicci, V. *et al.* Efficient allelic exchange and transposon mutagenesis in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **94**, 10955–10960 (1997).
- Mitchison, D. A. Treatment of tuberculosis. The Mitchell lecture 1979. *J. R. Coll. Physicians Lond.* **14**, 91–99 (1980).
- Girling, D. J. in *The Biology of the Mycobacteria, Volume 3: Clinical Aspects of Mycobacterial Disease* (eds Ratledge, C., Stanford, J. & Grange, J. M.) 285–323 (Academic, London, 1989).
- Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. Jr Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol. Microbiol.* **4**, 1911–1919 (1990).
- Stover, C. K. *et al.* New use of BCG for recombinant vaccines. *Nature* **351**, 456–460 (1991).
- Pavelka, M. S. & Jacobs, W. R. Comparison of the construction of unmarked deletion mutants in *Mycobacterium smegmatis*, *Mycobacterium bovis* bacillus Calmette-Guerin, and *Mycobacterium tuberculosis* H37Rv by allelic exchange. *J. Bacteriol.* **181**, 4780–4789 (1999).
- Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
- Russo-Marie, F., Roederer, M., Sager, B., Herzenberg, L. A. & Kaiser, D. Beta-galactosidase activity in single differentiating bacterial cells. *Proc Natl Acad Sci USA* **90**, 8194–8198 (1993).
- Sharma, V. *et al.* The structure of isocitrate lyase from *Mycobacterium tuberculosis*: a lynchpin to persistence of infection. *Nature Struct. Biol.* (in the press).

Acknowledgements

J.D.M. and colleagues generated the data shown in Figs 1, 2, and 5. K.H.Z.B. and colleagues generated the data shown in Figs 3 and 4. We thank B. R. Bloom for helpful discussions. J.D.M. was supported by a fellowship from the Helen Hay Whitney Foundation. J.D.M., E.J.M., W.T.C., K.H.Z.B., A.M., D.S. and D.G.R. were supported by a grant from Glaxo Wellcome. All authors were supported by funds from the US Public Health Service.

Correspondence and requests for materials should be addressed to D.G.R. (e-mail: dgr8@cornell.edu).

Progression of autoimmune diabetes driven by avidity maturation of a T-cell population

Abdelaziz Amrani*, Joan Verdaguera*, Pau Serra*, Sabrina Tafuro†, Rusung Tan‡ & Pere Santamaria*

* Department of Microbiology and Infectious Diseases, Faculty of Medicine, University of Calgary, Health Sciences Centre, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada

† MRC Human Immunology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

‡ Department of Pathology and Laboratory Medicine, University of British Columbia and BC's Children's Hospital, 4480 Oak Street, Vancouver, V6H 3V4, British Columbia, Canada

For unknown reasons, autoimmune diseases such as type 1 diabetes develop after prolonged periods of inflammation of mononuclear cells in target tissues¹. Here we show that progression of pancreatic islet inflammation to overt diabetes in non-obese diabetic (NOD) mice is driven by the 'avidity maturation' of a prevailing, pancreatic beta-cell-specific T-lymphocyte population carrying the CD8 antigen. This T-lymphocyte population recognizes two related peptides (NRP and NRP-A7)² in the context of H-2K^d class I molecules of the major histocompatibility complex (MHC). As pre-diabetic NOD mice age, their islet-associated CD8⁺ T lymphocytes contain increasing numbers of NRP-A7-reactive cells, and these cells bind NRP-A7/H-2K^d tetramers with increased specificity, increased avidity and longer half-lives. Repeated treatment of pre-diabetic NOD mice with soluble NRP-A7 peptide blunts the avidity maturation of the NRP-A7-reactive CD8⁺ T-cell population by selectively deleting those clonotypes expressing T-cell receptors with the highest affinity and lowest dissociation rates for peptide-MHC binding. This inhibits the local production of T cells that are cytotoxic to beta cells, and halts the progression from severe insulinitis to diabetes. We conclude that avidity maturation of pathogenic T-cell populations may be the key event in the progression of benign inflammation to overt disease in autoimmunity.

Nonobese diabetic mice develop inflammation of pancreatic islets (insulinitis) at 3 weeks of age, but do not begin to develop diabetes until 10 weeks later¹. Although CD8⁺ cytotoxic T lymphocytes (CTLs) are well established as killers of insulin-producing pancreatic beta cells in autoimmune diabetes^{3–9}, the mechanisms that drive the progression of insulinitis to overt diabetes remain a mystery. It has been shown that most of the islet-associated CD8⁺ CTLs in NOD mice are cytotoxic to beta cells in the context of the MHC class I molecule H-2K^d, and that many of these CTLs express highly homologous T-cell receptor (TCR) α -subunit chains (V α 17 and J α 42 elements joined by the N-region sequence M-R-D/E)^{7,9,10}. By screening combinatorial peptide libraries, we have identified two peptide ligands for this group of CTLs: NRP (amino-acid sequence KYNKANWFL), and NRP-A7 (KYNKANAFLL), an alanine mutant analogue of NRP with superior agonistic properties². As many of the CTL clones derived from islets of diabetic NOD mice recognize these peptides², and as transgenic NOD mice expressing an NRP/NRP-A7-reactive TCR (8.3-TCR $\alpha\beta$ -transgenic NOD mice) become diabetic shortly after developing insulinitis⁸, we thought that progression of insulinitis to diabetes might be driven by the accumulation of NRP/NRP-A7-reactive CTLs in islets.

To follow the accumulation of antigen-specific CD8⁺ T cells in pancreatic islets of pre-diabetic NOD mice, we generated H-2K^d tetramers containing the peptides NRP, NRP-A7, TUM (KYQAVTTTL, a negative control¹¹) or INS (LYLVCGERG, an

insulin-derived peptide recognized by islet-associated T cells from young NOD mice¹²). The NRP and NRP-A7 tetramers stained virtually all the splenic CD8⁺ T cells from 8.3-TCR $\alpha\beta$ -transgenic NOD mice, but less than 1% of the splenic CD8⁺ T cells from NOD mice. The INS tetramer stained the INS-reactive CTL clone G9C8 (ref. 12), but neither this tetramer nor the negative control tetramer (TUM) stained the splenic CD8⁺ T cells from NOD or 8.3-TCR $\alpha\beta$ -transgenic NOD mice (data not shown).

As tetramer-reactive cells represent a very small fraction of all the T cells contained in freshly isolated islets (see Fig. 1 legend), we assayed for tetramer-reactive clonotypes within the cell population that has undergone antigen-driven activation and thus is likely to contain autoreactive specificities. Islets from non-diabetic 5-, 9-, 15- and 20-week-old NOD mice were cultured in the presence of recombinant interleukin-2 (rIL-2) for 6–7 days, to selectively expand IL-2 receptor-positive T cells, which constitute about 10% of all the islet-associated CD8⁺ T cells⁷. There were age-dependent increases in the absolute number of T cells recovered from NOD

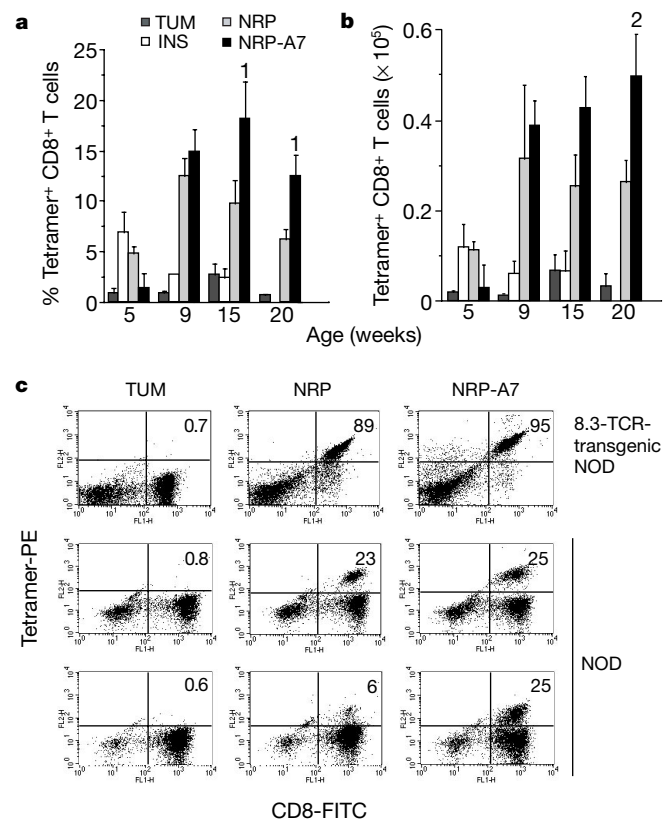


Figure 1 TUM-, INS-, NRP- and NRP-A7-reactive CD8⁺ T cells derived from pancreatic islets. **a**, Percentage of tetramer-reactive T cells. Data (mean $\pm$ s.e.m.) are from mice aged 5 weeks ($n = 3$), 9 weeks ($n = 13$), 15 weeks ($n = 7$) and 20 weeks ($n = 11$). *P* values: INS and NRP versus TUM at 5 weeks, $P < 0.04$; NRP and NRP-A7 versus TUM or INS at 9 weeks, $P < 0.0006$; NRP or NRP-A7 versus TUM or INS at 15 weeks, $P < 0.013$ and $P < 0.003$; NRP and NRP-A7 versus TUM at 20 weeks, $P < 0.0001$. '1', NRP-A7 versus NRP at 15 and 20 weeks, $P < 0.05$ and $P < 0.008$, respectively. Percentage of tetramer-reactive CD8⁺ cells in freshly isolated islets are $0.13 \pm 0.38\%$ ($n = 6$; 5 weeks), $0.95 \pm 0.67\%$ ($n = 4$; 9 weeks) and $1.6 \pm 0.5\%$ ($n = 2$; 15 weeks) (NRP-A7); and $0.71 \pm 0.4\%$ ($n = 4$; 5 weeks) and $0.31 \pm 0.4\%$ ($n = 4$; 9 weeks) (INS). **b**, Absolute number (mean $\pm$ s.e.m.) of tetramer-reactive T cells from the same samples as in **a**. *P* values: NRP-A7 at 15 and 20 weeks versus 9 and 5 weeks, $P < 0.04$; NRP and NRP-A7 versus TUM at 9 weeks, $P < 0.001$; NRP or NRP-A7 versus INS at 9 weeks, $P < 0.035$; NRP or NRP-A7 versus TUM or INS at 15 weeks, $P < 0.037$; NRP or NRP-A7 versus TUM at 20 weeks, $P < 0.0001$. '2', NRP-A7 versus NRP at 20 weeks, $P < 0.05$. **c**, Representative tetramer staining patterns. Over 90% of the CD8-negative cells shown in the plots are CD4⁺.

islets (from $2.3 \pm 0.2 \times 10^5$ cells at 5 weeks to $4.3 \pm 0.7 \times 10^5$ cells at 20 weeks, $P < 0.015$) and in the percentage of CD8⁺ T cells contained within islets (from $53 \pm 7\%$ at 5 weeks to $84 \pm 6\%$ at 20 weeks, $P < 0.003$). Tetramer staining of these cells indicated that NRP/NRP-A7-reactive CD8⁺ T cells constitute a significant fraction of the *in vivo* activated CD8⁺ T cells contained within islets (Fig. 1a, b). Notably, the size of the NRP-A7-reactive CD8⁺ T-cell population

increased with age, reaching an average of about 18% ($4.3 \pm 0.1 \times 10^4$ cells) by 15 weeks and about 13% ($5 \pm 0.1 \times 10^4$ cells) by 20 weeks (Fig. 1a, b). This increase was antigen-specific, as the number of NRP/NRP-A7-reactive CD8⁺ T cells was significantly greater than the number of INS-reactive (and TUM-reactive) CD8⁺ T cells at all ages tested, except at 5 weeks (Fig. 1a, b). Unexpectedly, the number of NRP-A7-reactive CD8⁺ T cells at 20 weeks was

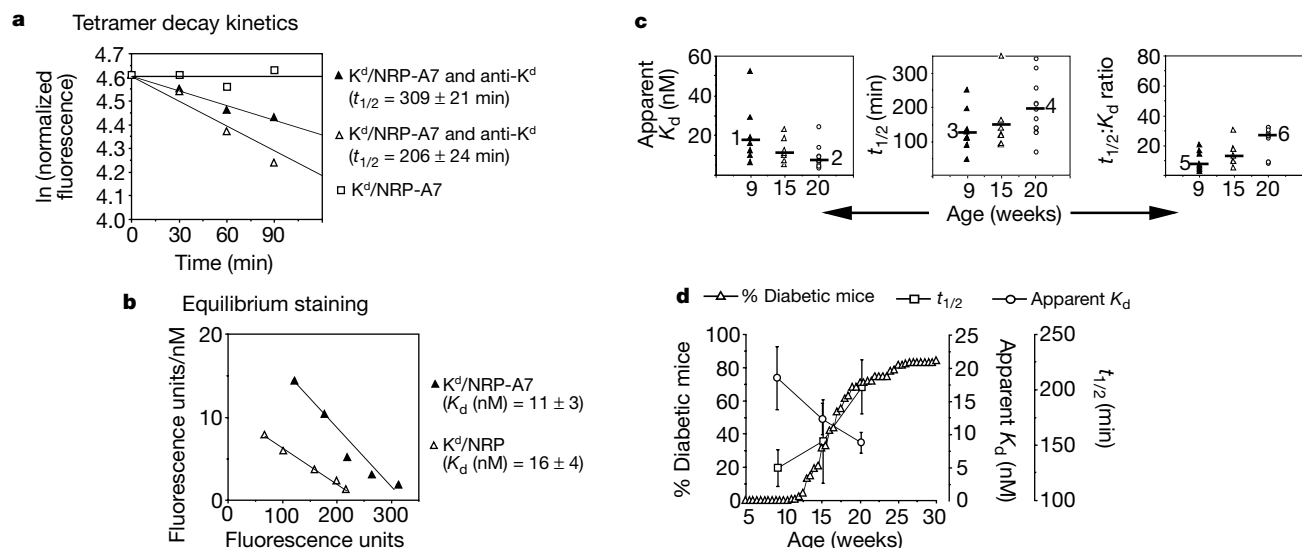


Figure 2 Association and dissociation kinetics of NRP and NRP-A7 tetramers. **a**, Tetramer decay kinetics of 8.3-TCR $\alpha\beta$ -transgenic CD8⁺ T cells ($n = 3$). **b**, Scatchard analysis of tetramer binding to 8.3-TCR $\alpha\beta$ -transgenic CD8⁺ T cells ($n = 3$). Cells were stained with different concentrations of NRP and NRP-A7 tetramers and the fluorescence units divided by the concentration (nM) (bound tetramer/free tetramer) plotted against fluorescence units (bound tetramer). At the same concentration, the NRP tetramer occupies fewer T-cell receptors on 8.3-CD8⁺ T cells than does the NRP-A7 tetramer. **c**, NRP-A7 tetramer association and dissociation kinetics for islet-derived CD8⁺ T cells from pre-diabetic NOD mice ($n = 9, 7$ and 13 for K_d ; $n = 11, 7$ and 12 for $t_{1/2}$; and $n = 8, 7$ and 12 for

K_d ; $t_{1/2}$ ratios at 9, 15 and 20 weeks, respectively). P values for cells from mice at 9 weeks versus 20 weeks: K_d , $P < 0.026$ ('1' versus '2'); $t_{1/2}$, $P < 0.027$ ('3' versus '4'); $t_{1/2}:K_d$ ratio, $P < 0.015$ ('5' versus '6'). The correlation coefficients for the average decay slopes (regression 0–90 min) of samples from mice at 9, 15 and 20 weeks were 0.907 ($P < 0.048$); 0.902 ($P = 0.054$); 0.993 ($P < 0.004$). There were statistically significant differences for the average of the logarithm of the normalized fluorescence at 90 min between 9-week-old and 20-week-old groups (4.03 ± 0.08 versus 4.29 ± 0.08 , $P < 0.013$). **d**, 'Avidity maturation' of NRP-A7-reactive cells (mean $\pm$ s.e.m.) versus diabetes penetrance.

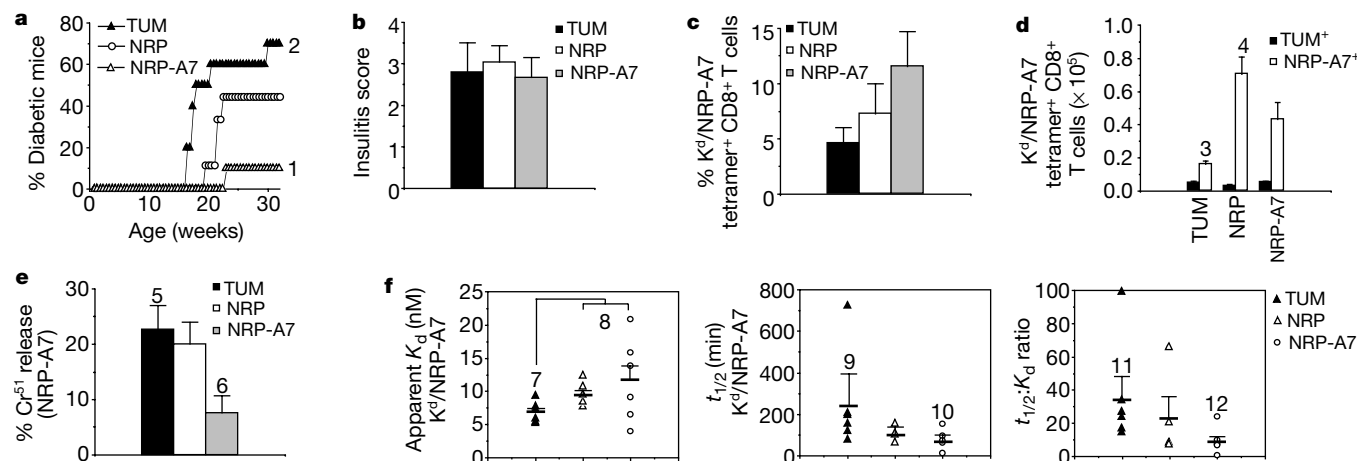


Figure 3 Diabetogenesis in peptide-treated NOD mice. **a**, Incidence of diabetes in groups of 10 female NOD mice treated with i.p. injections of TUM, NRP and NRP-A7 peptides in PBS. '1' versus '2', $P < 0.007$. **b**, Insulinitis scores in non-diabetic mice ($n = 2$ –3 mice per group; mean $\pm$ s.e.m.). **c**, **d**, Percentage (**c**) and absolute number (**d**) of NRP-A7 tetramer-reactive CD8⁺ cells in peptide-treated (i.p. and f.p.) mice ($n = 8$ NRP-A7, 9 NRP and 10 control mice). '3' versus '4', $P < 0.015$. Percentage and number of CD8⁺ cells isolated from peptide-treated mice were $74 \pm 3\%$ and $4 \pm 0.2 \times 10^5$ (TUM); $90 \pm 1\%$ and $5.7 \pm 1 \times 10^5$ (NRP); and $86 \pm 3\%$ and $5.3 \pm 0.9 \times 10^5$ (NRP-A7) ($P < 0.03$ for NRP- plus NRP-A7-treated versus control mice). No differences between diabetic and non-diabetic mice were observed. **e**, NRP-A7-specific minus TUM-specific cytotoxicity of

CD8⁺ T cells from peptide-treated (i.p. and f.p.) mice ($n = 7$ NRP-A7, 11 NRP and 13 control mice; mean $\pm$ s.e.m.). No differences between diabetic and non-diabetic mice were noted. '5' versus '6', $P < 0.017$. **f**, NRP-A7 tetramer binding kinetics of CD8⁺ T cells from non-diabetic peptide-treated (i.p. and f.p.) mice at 32 weeks ($n = 5$ NRP-A7, 4 NRP and 6 control mice; mean $\pm$ s.e.m.). '7' versus '8', $P < 0.05$; '9' versus '10', $P < 0.04$; '11' versus '12', $P < 0.02$. The correlation coefficients for the average decay slopes (0–90 min) of samples from control peptide-, NRP- and NRP-A7-treated mice were 0.954 ($P < 0.023$); 0.899 ($P = 0.054$); 0.979 ($P < 0.03$). The ln (normalized fluorescence) values at 90 min between the control peptide- and NRP-A7-treated groups were 3.88 ± 0.12 versus 4.31 ± 0.15 ($P = 0.055$).

significantly greater than the number of NRP-reactive CD8⁺ T cells (Fig. 1a, b).

Because all NRP-reactive lines were also reactive to NRP-A7, but not *vice versa* (Fig. 1c), and at least one of the V α 17-MRD-J α 42-expressing T-cell clones⁹ recognizes NRP-A7 but not NRP (T. DiLorenzo, D. Serreze, P. S. and S. Nathenson, unpublished data), these data suggested that the age-dependent increase in the size of the NRP/NRP-A7-reactive population was accompanied by an outgrowth of clonotypes that are capable of recognizing NRP-A7 but not NRP. Furthermore, because NRP/NRP-A7-reactive clonotypes (that is, 8.3-TCR $\alpha\beta$ -transgenic CTLs) bind the NRP-A7 tetramer with a longer half-life and higher avidity (lower K_d) than the NRP tetramer (Fig. 2a, b), we reasoned that the outgrowing clonotypes might be those with the highest avidity for NRP-A7/H-2K^d binding.

To test this hypothesis, we measured the association (K_d) and dissociation kinetics ($t_{1/2}$)¹³ of the interaction between NRP-A7 tetramers and islet-derived CD8⁺ T cells obtained at different time points (9, 15 and 20 weeks). CD8⁺ T cells isolated from 20-week-old mice bound the NRP-A7 tetramer with significantly higher avidity and a longer half-life than CD8⁺ T cells isolated from 9-week-old mice (Fig. 2c). These kinetic changes could not be accounted for by changes in the expression levels of either the T-cell receptor (TCR) or CD8. The islet-associated NRP-A7-reactive CD8⁺ T-cell population therefore increases its overall avidity for peptide-MHC as the mice approach the age at which the incidence of diabetes within the colony follows an exponential distribution (Fig. 2d).

To determine whether development of diabetes in NOD mice requires the accumulation of NRP-A7-reactive CTLs in islets, we treated cohorts of female NOD mice with repeated intraperitoneal injections of 100 μ g of TUM, NRP or NRP-A7 in PBS. Previous studies in 8.3-TCR $\alpha\beta$ -transgenic NOD mice had indicated that repeated administration of NRP or NRP-A7 through intraperitoneal and footpad routes induced the deletion of CD8⁺ T-cells binding NRP and NRP-A7 tetramers with high avidity and slow dissociation kinetics (our unpublished data). TUM-treated NOD mice developed diabetes with an incidence similar to untreated NOD mice (Fig. 3a; see Fig. 2d). Similar results were obtained using a cohort of mice treated with another negative control peptide (LCMV-GP33) or with the INS peptide (data not shown). NOD mice treated with NRP developed diabetes at a slightly lower rate when compared with TUM-treated mice (Fig. 3a). In contrast, mice treated with NRP-A7 were highly protected from diabetes (Fig. 3a). Administration of NRP-A7 through the footpad was also protective ($P < 0.04$ when compared with TUM-treated mice; data not shown). These results therefore indicated that NRP-A7-reactive CD8⁺ T cells have a critical role in diabetogenesis.

To ascertain whether the NRP-A7 (but not NRP) peptide treatment had induced the deletion of NRP-A7-reactive CD8⁺ T cells, we examined islets of non-diabetic 32-week-old mice from each treatment group for the presence of this T-cell population. Immunopathological studies revealed that TUM-, NRP- and NRP-A7-treated non-diabetic mice had similar insulinitis scores (Fig. 3b), and that their insulinitis lesions contained similar proportions of CD4⁺ and CD8⁺ T cells (data not shown). Unexpectedly, the islet-derived CD8⁺ T cells of NRP-A7- and NRP-treated mice contained similar, if not greater numbers of NRP-A7 tetramer-reactive cells than did control peptide-treated mice (Fig. 3c, d), refuting the idea that the NRP-A7-reactive T-cell population had undergone massive deletion in NRP-A7-treated mice. However, the CD8⁺ T cells derived from NRP-A7-treated mice were significantly less cytotoxic to NRP-A7-pulsed, H-2K^d complementary-DNA-transfected RMA-S (RMA-S-K^d) targets than those derived from NRP- or control peptide-treated mice (Fig. 3e), indicating that NRP-A7 treatment had somehow impaired the cytotoxic potential of the NRP-A7-reactive T-cell population.

To determine whether the reduced cytotoxic activity of these T cells resulted from the absence of clonotypes bearing high-affinity

TCRs for NRP-A7, we compared the kinetics of NRP-A7 tetramer binding to CD8⁺ T cells derived from NRP-, NRP-A7- and control peptide-treated mice. CD8⁺ T cells derived from NRP-A7-treated mice bound the NRP-A7 tetramer with a lower avidity and shorter half-life than CD8⁺ T cells derived from NRP- or control peptide-treated mice (Fig. 3f). These results show that treatment with NRP-A7 selectively eliminated those NRP-A7-reactive CD8⁺ T cells that expressed high-affinity TCRs for NRP-A7, while allowing the local expansion of clonotypes bearing lower-affinity TCRs. As treatment with NRP-A7 also halted the progression from benign to pernicious islet inflammation, we conclude that development of overt diabetes in NOD mice requires the affinity maturation of the NRP-A7-reactive CD8⁺ T-cell population to avidity levels above a diabetogenic threshold.

Our results provide new insight into the pathogenesis of autoimmunity. Autoimmune diseases often develop after prolonged periods of target organ inflammation that produce limited tissue damage. In human autoimmune diabetes, for example, siblings of diabetic probands develop signs of beta-cell autoreactivity, in the form of circulating autoantibodies and impaired beta-cell function, many years before they develop overt disease¹⁴. A similar phenomenon occurs in the NOD mouse¹. Our data provide an explanation for the puzzling resistance of young NOD mice (and of a percentage of older NOD mice with severely inflamed pancreatic islets) to the development of diabetes. The progression from benign to destructive tissue inflammation in autoimmune diabetes is a slow and incompletely penetrant process because it requires the avidity maturation of at least one highly pathogenic autoreactive CD8⁺ T-cell population. In the NOD mouse, this population of cells is represented by clonotypes that share similar TCR α chains^{7,9,10} and that recognize the peptide NRP-A7 in the context of H-2K^d molecules².

We propose that, after the recruitment of putative disease-initiating T cells to islets (such as those recognizing insulin or glutamic acid decarboxylase^{15–17}), NRP-A7-specific CD8⁺ clonotypes expressing high-affinity TCRs undergo preferential expansion *in situ*, at the expense of clonotypes bearing low-affinity TCRs. This phenomenon is likely to be analogous to the *in vivo* maturation process undergone by antigen-specific CD4⁺ T cells in the regional lymph nodes after a secondary antigenic challenge, a process that promotes the selection of CD4⁺ T cells capable of binding peptide-MHC class II complexes with higher avidity and/or slower dissociation kinetics¹³. As the biological effects of TCR-antigen-MHC interactions, including TCR oligomerization, aggregation, cluster formation and signalling, often correlate with their avidity and duration^{18–23}, this maturation process probably selects for T cells with increasing cytotoxic and diabetogenic potential. By killing beta cells efficiently, these CTLs would release sequestered beta cell autoantigens into the local milieu, promoting the avidity maturation of additional autoreactive T-cell specificities and the amplification of a self-perpetuating destructive process that results in the complete destruction of pancreatic beta cells. □

Methods

Mice

The 8.3-TCR $\alpha\beta$ -transgenic NOD mice have been described⁸. NOD mice were purchased from Taconic Farms.

Peptides and tetramers

The peptides TUM, INS, NRP and NRP-A7 were prepared using FMOC chemistry, purified by reverse phase high-performance liquid chromatography to more than 80% purity, and sequenced by ion spray mass spectrometry (Chiron Technologies). Tetramers were prepared as described²⁴.

Islet-associated CD8⁺ T cells

Pancreatic islets from NOD mice were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum and 0.5 U ml⁻¹ of Takeda rIL-2 for 6–7 days.

Tetramer binding kinetics and flow cytometry

Kinetic analyses of tetramer binding were conducted as described¹³, with minor mod-

ifications. For the analysis of tetramer staining at equilibrium, T cells were stained for 2 h at room temperature with 20 μ l of wash medium (0.2% sodium bicarbonate, 0.1% sodium azide and 2% FBS in RPMI-1640) containing anti-CD8 α -FITC (clone YTS169.4; 0.5 μ g) and different concentrations of tetramers (8.55, 17.1, 42.75 and 85.5 nM). Titration studies with 8.3-TCR $\alpha\beta$ -transgenic CD8 $^+$ T cells indicated that tetramer staining in the presence of this anti-CD8 monoclonal antibody does not change the percentage or absolute number of cells that bind the tetramer. After washing, cells were resuspended in 100 μ l of wash medium and analysed with a flow cytometer.

The apparent K_d values were determined by plotting the negative reciprocal of the slope of the line fit to Scatchard plots of fluorescence units (median of CD8 $^+$ population staining for tetramer) divided by concentration (nM) versus fluorescence units. For the analysis of binding half-life, T cells were stained with 85.5 nM tetramer and 0.5 μ g of anti-CD8-FITC in 20 μ l of wash medium for 45 min at room temperature. Cells were washed three times, cooled to 4°C and resuspended in 100 μ l of cold medium containing 50 μ g of an anti-K d monoclonal antibody (20–8–4S) or no antibody at all (negative control). Aliquots of 20 μ l were taken at 0, 30, 60 and 90 min, diluted in 200 μ l of wash medium and analysed with a flow cytometer. Half-lives ($t_{1/2}$) were determined by calculating the $(\ln 2)/\text{mean slope value}$ of plots of the corrected natural logarithm ($\ln$) of the percentage normalized fluorescence (sum of fluorescence intensities of tetramer-positive CD8 $^+$ T cells normalized per CD8 $^+$ T cell at each point with respect to the initial time point) versus time. Studies of binding decay using samples from 9-week-old NOD mice indicated that half-lives calculated from 0–120-min slopes were not statistically different from the half-life values calculated from 0–90-min slopes ($n = 6$; 138 ± 17 versus 165 ± 24 min, respectively). The 120-min analysis was therefore omitted from studies at older ages, to overcome limitations in cell numbers.

⁵¹Cr-release assays

Cytotoxicity assays were done as described², using peptide-pulsed (1 μ g ml $^{-1}$) RMA-S-K d cells labelled with [51 Cr]sodium chromate as targets at a 1:10 target:effector ratio.

Peptide treatment

Three week-old female NOD mice were injected with 100 μ g of peptide in PBS intraperitoneally (i.p.) or through the footpad (f.p.). This was done every 2 weeks until week 7, and every 3 weeks thereafter. Mice were monitored for development of hyperglycaemia until week 32.

Pathology

Formalin-fixed, paraffin-embedded pancreas sections were stained with haematoxylin and eosin and scored for insulinitis as described⁸.

Statistical analyses

Data were compared using linear regression and variance analysis, Mann–Whitney *U* test or χ^2 test.

Received 14 April; accepted 8 June 2000.

- Delovitch, T. & Singh, B. The nonobese diabetic mouse as a model of autoimmune diabetes: immune dysregulation gets the NOD. *Immunity* **7**, 727–738 (1997).
- Anderson, B., Park, B. J., Verdager, J., Amrani, A. & Santamaria, P. Prevalent CD8 $^+$ T cell response against one peptide/MHC complex in autoimmune diabetes. *Proc. Natl Acad. Sci. USA* **96**, 9311–9316 (1999).
- Nagata, M., Santamaria, P., Kawamura, T., Utsugi, T. & Yoon, J.-W. Evidence for the role of CD8 $^+$ cytotoxic T cells in the destruction of pancreatic beta cells in NOD mice. *J. Immunol.* **152**, 2042–2050 (1994).
- Katz, J., Benoist, C. & Mathis, D. Major histocompatibility complex class I molecules are required for the generation of insulinitis in non-obese diabetic mice. *Eur. J. Immunol.* **23**, 3358–3360 (1993).
- Wicker, L. *et al.* β 2-microglobulin-deficient NOD mice do not develop insulinitis or diabetes. *Diabetes* **43**, 500–504 (1994).
- Serreze, D., Leiter, E., Christianson, G., Greiner, D. & Roopenian, D. Major histocompatibility complex class I-deficient NOD.B1m mll mice are diabetes and insulinitis resistant. *Diabetes* **43**, 505–508 (1994).
- Verdager, J. *et al.* Acceleration of spontaneous diabetes in TCR β -transgenic nonobese diabetic mice by beta cell-cytotoxic CD8 $^+$ T cells expressing identical endogenous TCR α chains. *J. Immunol.* **157**, 4726–4735 (1996).
- Verdager, J. *et al.* Spontaneous autoimmune diabetes in monoclonal T cell nonobese diabetic mice. *J. Exp. Med.* **186**, 1663–1676 (1997).
- DiLorenzo, T. *et al.* Major histocompatibility complex class I-restricted T cells are required for all but the end stages of diabetes development in nonobese diabetic mice and use prevalent T cell receptor α chain gene rearrangement. *Proc. Natl Acad. Sci. USA* **95**, 12538–12543 (1998).
- Santamaria, P. *et al.* Beta cell cytotoxic CD8 $^+$ T cells from non-obese diabetic mice use highly homologous T cell receptor alpha chain CDR3 sequences. *J. Immunol.* **154**, 2494–2503 (1995).
- Rammensee, H. G., Friede, T. & Stevanović, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* **41**, 178–228 (1995).
- Wong, S. *et al.* Identification of an MHC class I-restricted autoantigen in type 1 diabetes by screening an organ-specific cDNA library. *Nature Med.* **5**, 1026–1031 (1999).
- Savage, P. A., Boniface, J. J. & Davis, M. M. A kinetic basis for T cell receptor repertoire selection during an immune response. *Immunity* **10**, 485–492 (1999).
- Abiru, N. & Eisenbarth, G. Autoantibodies and autoantigens in type 1 diabetes: role in pathogenesis, prediction and prevention. *Can. J. Diab. Care* **23**, 59–65 (1999).
- Kaufman, D. *et al.* Spontaneous loss of T-cell tolerance to glutamic acid decarboxylase in murine insulin-dependent diabetes. *Nature* **366**, 69–72 (1993).
- Tisch, R. *et al.* Immune response to glutamic acid decarboxylase correlates with insulinitis in non-obese diabetic mice. *Nature* **366**, 72–75 (1993).
- Wegmann, D. The immune response to islet in experimental diabetes and insulin-dependent diabetes mellitus. *Curr. Opin. Immunol.* **8**, 860–864 (1996).
- Alam, S. *et al.* T-cell receptor affinity and thymocyte positive selection. *Nature* **381**, 616–620 (1996).

- Kersh, G., Kersh, E., Fremont, D. & Allen, P. High and low potency ligands with similar affinities for the TCR: the importance of kinetics in TCR signaling. *Immunity* **9**, 817–826 (1998).
- Lyons, D. *et al.* A TCR binds to antagonist ligands with lower affinities and faster dissociation rates than to agonists. *Immunity* **5**, 53–61 (1996).
- Matsui, K., Boniface, J., Steffner, P., Reay, P. & Davis, M. Kinetics of T-cell receptor binding to peptide/I-A k complexes: correlation of the dissociation rate with T-cell responsiveness. *Proc. Natl Acad. Sci. USA* **91**, 12862–12866 (1994).
- Monks, C., Freiberg, B., Kupfer, H., Sciaky, N. & Kupfer, A. Three dimensional segregation of supramolecular activation clusters in T-cells. *Nature* **395**, 82–86 (1998).
- Reich, Z. *et al.* Ligand-specific oligomerization of T-cell receptor molecules. *Nature* **387**, 617–620 (1997).
- Altman, J. *et al.* Direct visualization and phenotypic analysis of virus-specific T lymphocytes in HIV-infected individuals. *Science* **274**, 94–96 (1996).

Acknowledgements

The authors thank S. Bou, K. Rouleau, M. Deumas and S. Culp for animal care; S. Thiessen for scoring insulinitis; M. Bevan, K. Kane and T. Utsugi for reagents; S. Wong for testing the INS tetramer; Y. Yang for reading the manuscript; and H. Kominck for editorial assistance. This work was supported by grants from the Medical Research Council of Canada, the MRC/Juvenile Diabetes Foundation International (JDFI) Program II and the Canadian Diabetes Association (to P.S.). A.A. was supported by fellowships from the Alberta Heritage Foundation for Medical Research (AHFMR) and the JDFI. P.S. is a Senior Scholar of the AHFMR. R.T. and P.S. are members of the Beta Cell Apoptosis Network (BetaCAN).

Correspondence and requests for materials should be addressed to P.S. (e-mail: psantama@ucalgary.ca).

The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells

Peter J. P. Coopman[†], Michael T. H. Do[†], Mara Barth[†], Emma T. Bowden[‡], Andrew J. Hayes[‡], Eugenia Basyuk[§], Jan K. Blancato^{||}, Phyllis R. Vezza[¶], Sandra W. McLeskey[‡], Paul H. Manges[‡] & Susette C. Mueller[†]

Departments of ^{*} Cell Biology, [‡] Oncology, [¶] Pathology, [§] Pharmacology and School of Nursing, ^{||} Institute for Molecular and Human Genetics, and Vincent T. Lombardi Cancer Center, Georgetown University Medical School, Washington DC 20007, USA

[§] Centre National de la Recherche Scientifique UPR 9023 and [†] Centre National de la Recherche Scientifique UMR 5539, 34095 Montpellier, France

Syk is a protein tyrosine kinase that is widely expressed in haematopoietic cells. It is involved in coupling activated immunoreceptors to downstream signalling events that mediate diverse cellular responses including proliferation, differentiation and phagocytosis^{1–4}. Syk expression has been reported in cell lines of epithelial origin⁵, but its function in these cells remains unknown. Here we show that Syk is commonly expressed in normal human breast tissue, benign breast lesions and low-tumorigenic breast cancer cell lines. Syk messenger RNA and protein, however, are low or undetectable in invasive breast carcinoma tissue and cell lines. Transfection of wild-type Syk into a Syk-negative breast cancer cell line markedly inhibited its tumour growth and metastasis formation in athymic mice. Conversely, overexpression of a kinase-deficient Syk in a Syk-positive breast cancer cell line significantly increased its tumour incidence and growth. Suppression of tumour growth by the reintroduction of Syk appeared to be the result of aberrant mitosis and cytokinesis. We propose that Syk is a potent modulator of epithelial cell growth and a potential tumour suppressor in human breast carcinomas.

To determine whether Syk has a role in breast cancer phagocytosis and invasion^{6,7}, we tested the expression of Syk in a panel of well-

[†] Present address: Centre National de la Recherche Scientifique UMR 5539, Montpellier, France.

characterized human breast cancer cell lines by polymerase chain reaction with reverse transcription (RT-PCR) (Fig. 1a, b). K562 erythroleukaemic cells, known to express Syk, were used as a positive control. Syk was expressed in normal human mammary gland tissue as well as in the human breast epithelium cell line MCF10A. It was also strongly expressed in several human breast carcinoma cell lines (MDA-MB-468, BT474, MCF7, SKBr3), but poorly expressed (Hs578T, MDA-MB-436) or undetectable in others (BT549, MDA-MB-231, MDA-MB-435). Nucleotide sequencing and Southern hybridization with an internal Syk probe (Fig. 1a, Syk-Biot.) proved that we had amplified Syk fragments. We detected Syk protein by immunoprecipitation and western blotting of cell extracts (Fig. 1c). Syk protein levels in the various breast cancer cell lines were comparable with those in the K562 cells. Corresponding to the absence of Syk mRNA, several breast cancer cell lines did not show any detectable Syk protein. Most strikingly, Syk was absent only in highly tumorigenic breast cancer cell lines displaying typical features of a malignant phenotype including increased motility and invasion⁸ (Table 1). The ZAP-70 protein tyrosine kinase, which shares high amino-acid sequence homology with Syk⁹, was not detected in the breast cancer cell lines (data not shown). These observations indicate that loss of Syk expression may be associated with the progression towards a malignant phenotype in breast cancer.

The Syk protein expressed by breast tumour cells was catalytically active (Fig. 1d). *In vitro*, Syk from MCF7 cells was autophosphorylated in an immunocomplex kinase assay. Treatment of cells with pervanadate, a phosphotyrosine phosphatase inhibitor, led to activation of Syk¹⁰ and induced Syk autophosphorylation in

Table 1 Syk protein and mRNA expression in breast cancer cell lines

Breast cancer cell line	<i>In vivo</i> (nu/nu)*	Chemo-invasion†	Matrigel morphology‡	Syk RNA§	Syk protein
MDA-MB-231	LI	+++++	INV	–	–
BT549	N	+++++	INV	–	–
Hs578T	HM	+++++	INV	+/-	–
MCF7-ADR	P	+++	INV	n.d.	–
MDA-MB-435	LI	+++	INV	–	–
MDA-MB-436	LI	+++	INV	+/-	–
MCF7	P	++	SPH	+	+
BT474	P	+	SPH	+	+
MDA-MB-468	P	+	SPH	+	+
SKBr3	N	+	SPH	+	+
MCF10A	nd	nd	nd	+	+

*Activity in athymic nude mice: N, non-tumorigenic; P, primary tumour formation, no local invasiveness or metastasis; LI = local invasion through peritoneum, colonization of visceral organs; HM, hematogenous metastasis to the lungs; n.d., not determined. Data taken from ref. 8. †Activity in Boyden chamber chemoinvasion assay. Data taken from ref. 8. ‡Morphology in Matrigel: SPH, spherical colony or non-invasive cluster formation; INV, invasive colony formation. Data taken from ref. 8. §As determined by RT-PCR (see Fig. 1a, b). ||As determined by immunoprecipitation followed by western blotting (see Fig. 1c).

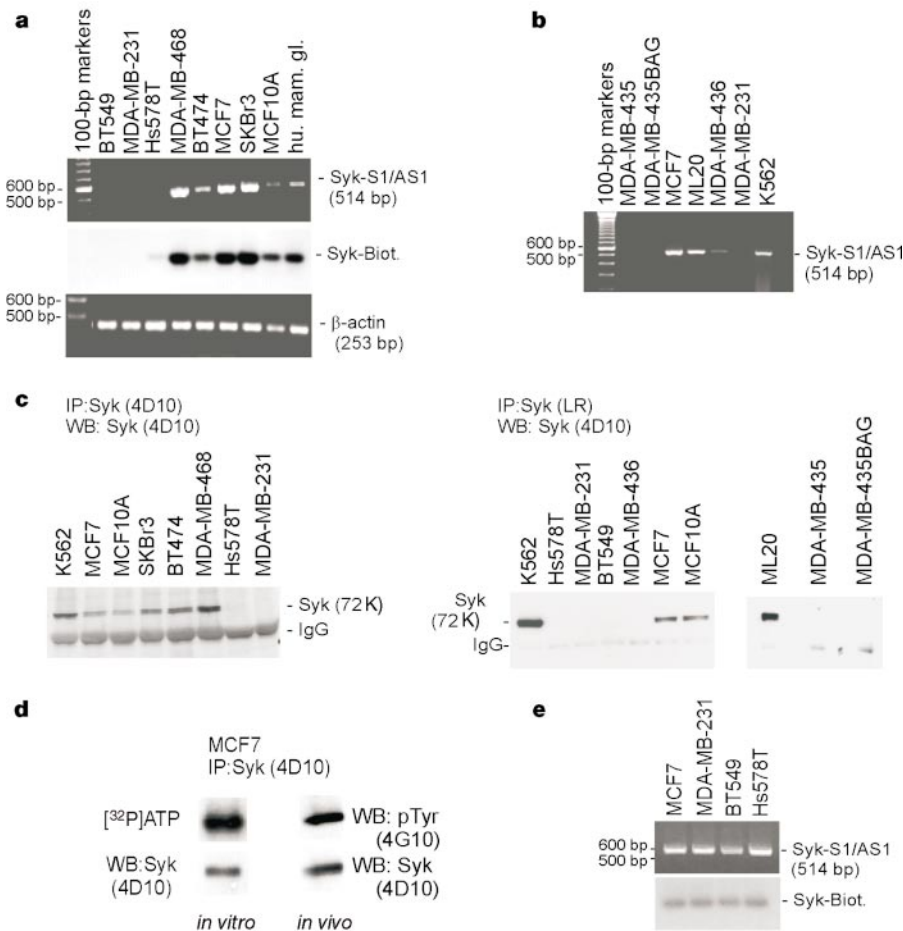


Figure 1 Syk expression in human breast cancer cell lines. **a**, RT-PCR detection of Syk and β -actin mRNA. Southern blotting of the RT-PCR products using a biotinylated internal Syk oligonucleotide (Syk-Biot.). hu. mam. gl., human mammary gland. **b**, RT-PCR detection of Syk mRNA in cell lines including the transfection acceptor cell lines ML20 and MDA-MB-435BAG and the K562 erythroleukaemia cell line. **c**, Immuno-

precipitation (IP) and western blot (WB) detection of Syk protein using polyclonal (LR) and monoclonal (4D10) antibodies. **d**, Autophosphorylation activity of immunoprecipitated Syk as detected by *in vitro* incorporation of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ or by western blotting with anti-phosphotyrosine (pTyr) antibodies. **e**, PCR and Southern blot detection of Syk in genomic DNA.

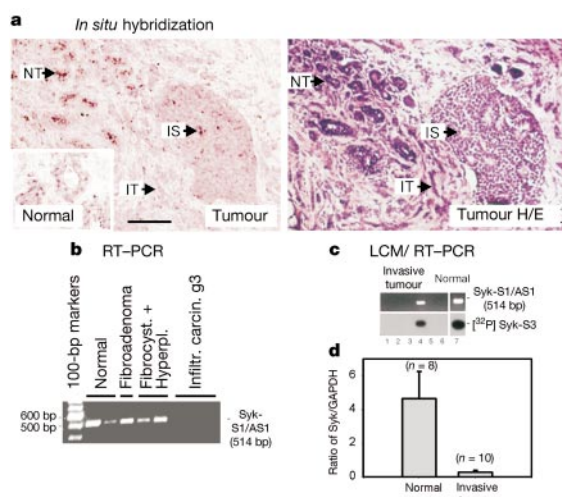


Figure 2 Syk expression in normal and pathological human breast tissue samples. **a**, *In situ* hybridization of Syk in normal breast tissue (NT), *in situ* tumour (IS) and invasive breast tumour (IT). H/E, haematoxylin and eosin stain. Scale bar, 500 μ m; inset is enlarged twofold. **b**, RT-PCR detection of Syk mRNA in normal breast tissue, benign breast disease and grade 3 infiltrating breast carcinoma. **c**, RT-PCR detection of Syk in invasive breast carcinoma and normal breast tissue isolated by laser capture microdissection (LCM) (Syk-S1/AS1). Southern blotting of the RT-PCR products using an internal oligonucleotide (Syk-S3/ 32 P). **d**, [32 P]Syk-S3/ 32 P]GAPDH Southern blot signal ratios from LCM/RT-PCR breast tumour samples (Mean values $\pm$ s.e.).

MCF7 cells. The absence of Syk protein in malignant breast cancer cell lines does not seem to be caused by the previously reported frameshift mutation that generates a premature termination codon¹¹ as we did not detect Syk mRNA in these cells. PCR of genomic DNA demonstrated that the Syk gene is still present in all tested cell lines regardless of their Syk expression (Fig. 1e). Together, our results suggest that regulation of the Syk expression in breast cancer cell lines occurs at the transcriptional level; however, the exact mechanism of regulation is unknown.

As Syk expression is lost in malignant human breast cancer cell lines, we extended our study to normal and pathological human breast tissue samples. *In situ* hybridization verified that Syk expression in breast epithelial cells aligned within the ducts, whereas Syk expression was reduced in *in situ* carcinoma and lost in invasive breast carcinoma cells (Fig. 2a). Using RT-PCR, we detected Syk mRNA in normal breast tissue and in benign breast lesions (fibroadenoma, fibrocystic disease with hyperplasia), but not in malignant breast cancer (grade 3 infiltrating carcinoma) (Fig. 2b). Breast tumours are histologically and biochemically heterogeneous and might also contain normal breast epithelial cells or tumour-infiltrating lymphocytes that express Syk. To analyse Syk expression in invasive breast carcinoma cells or adjacent normal breast tissue exclusively, we used laser capture microdissection (LCM) followed by RT-PCR. Unlike normal breast tissue, most invasive tumour samples did not show any detectable Syk expression (Fig. 2c). Furthermore, Syk/GAPDH ratios in the LCM samples were significantly higher ($P < 0.01$, Student's *t*-test) in normal breast

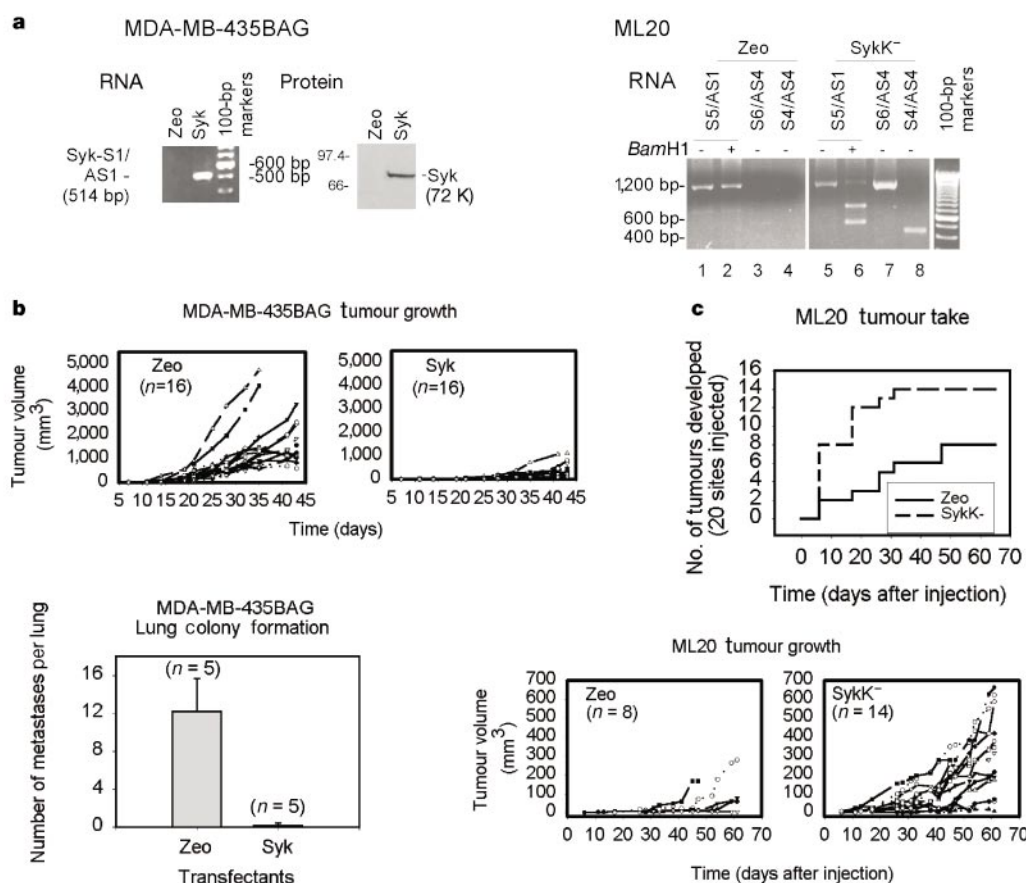


Figure 3 Effect of Syk transfection on tumorigenesis and metastasis formation *in vivo*. **a**, RT-PCR and western blot detection of wild-type Syk and kinase-negative Syk (SykK-) expression in MDA-MB-435BAG and ML20 stable transfectants. The S4, S6 and AS4 primers specifically amplify mutant porcine Syk. The S5 and AS1 primers amplify both human and porcine Syk, but *Bam*H1 digests only mutated porcine Syk. **b**, Tumour growth

and lung colony formation of wild-type Syk-transfected MDA-MB-435BAG cells in athymic mice. Lines represent growth of individual tumours. **c**, Tumour take and tumour growth of kinase-negative Syk-transfected ML20 cells in athymic ovariectomized mice, supplemented with a 17- β -oestradiol pellet. Lines represent growth of individual tumours.

epithelial cells as compared with invasive breast carcinoma cells (Fig. 2c). Our observations of cell lines and human tissue samples strongly suggest that Syk might block uncontrolled cell growth in normal breast tissue and that its absence may be permissive for malignant tumour growth.

We directly tested the effect of Syk transfection on tumorigenicity in an animal model. The use of *lacZ*-expressing acceptor cell lines allowed chromogenic detection of the transfected cells after injection in athymic mice. The Syk-negative MDA-MB-435BAG cells¹² were stably transfected with wild-type Syk complementary DNA. We verified Syk expression at both the mRNA (RT-PCR) and protein (western blotting) levels (Fig. 3a, MDA-MB-435BAG), and found that it was comparable to the endogenous levels in MCF7 cells. The catalytic activity of the transfected Syk was validated by its *in vitro* autophosphorylation activity (data not shown). Conversely,

Syk-positive ML20 cells¹³, a variant of the MCF7 cell line, were transfected with a kinase-deficient mutant of porcine Syk (SykK⁻) known to exert a dominant-negative function¹⁴. Its expression was specifically discriminated from the wild-type endogenous human Syk by RT-PCR using porcine-specific Syk primers (Fig. 3a, ML20, lanes 7–8). Less stringent primers amplified both human and porcine Syk, but a *Bam*H1 digest allowed specific detection of the point mutation introduced into the kinase-deficient porcine Syk. This approach showed that the dominant-negative Syk is expressed at higher levels than the wild-type Syk in ML20/SykK⁻ transfectants (Fig. 3a, ML20, lane 6, compare faint upper band with lower two bands). Expression of the kinase-deficient Syk did not, however, alter the protein levels of the endogenous wild-type Syk in ML20/SykK⁻ transfectants (data not shown).

Immediately after selection in zeocin, we injected the pooled transfected clones into the mammary fat pad of female athymic mice. All MDA-MB-435BAG tumours, regardless of the transfection, were already palpable 7 days after injection. Subsequently, the tumours of the MDA-MB-435BAG/Zeocin control transfectants grew steadily, attaining mean volumes of $2,004 \pm 1,158 \text{ mm}^3$ (mean $\pm$ s.d.) after 35 to 43 days (Fig. 3b). In contrast, the tumour growth of the Syk transfectants was significantly slower than that of the control transfectants ($P < 0.005$, Fisher variance analysis). At sacrifice (day 43), the Syk tumours reached a mean volume of only $419 \pm 297 \text{ mm}^3$, which was significantly smaller than the control (Zeocin) tumours ($P < 0.001$, Student's *t*-test). Mammary-fat-pad-injected MDA-MB-435BAG transfectants did not develop any detectable spontaneous metastases after 43 days. We therefore tested their capacity to form experimental lung metastases after injection in the tail vein. With the Syk transfectants, only one of five animals showed a single metastatic colony in the lungs (Fig. 3b). In contrast, all animals injected with control (Zeocin) cells developed several metastatic lung colonies. These results clearly show that re-expression of Syk by MDA-MB-435BAG cells was sufficient to suppress their tumour growth and metastasis formation. The latency period for tumour appearance of the mammary-fat-pad-injected ML20 transfectants was very variable and ranged from 6 to 47 days, regardless of the transfection (Fig. 3c). The potency of the SykK⁻ transfectants to develop a tumour (tumour take), however, was significantly increased as compared with the control (Zeocin) transfectants ($P < 0.001$, Student's *t*-test). In addition, the overall tumour growth of the SykK⁻ transfectants was significantly faster than that of the control transfectants ($P < 0.005$, Fisher variance analysis) (Fig. 3c). We conclude that expression of the dominant-negative, kinase-deficient Syk releases growth inhibition in MCF7 cells that is dependent upon Syk tyrosine kinase activity.

To investigate the mechanisms underlying the suppressive effect of Syk expression on tumour growth, we studied apoptosis by *in situ* TUNEL (terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick end labelling) staining of histological sections of the primary tumours of MDA-MB-435BAG transfectants. The number of apoptotic nuclei per square micrometre was determined for tumour sections exclusive of necrotic areas, 6.38 ± 6.58 (Zeocin, $n = 11$) and 5.49 ± 2.32 (Syk, $n = 7$), and exhibited no significant statistical difference (Student's *t*-test). Therefore, programmed cell death does not seem to account for the inhibited tumour growth; however, staining of the MDA-MB-435BAG tumour sections with haematoxylin and eosin highlighted abnormalities in cell division. Unlike control (Zeocin) tumours, Syk tumours exhibited groups of three to many cells that were enlarged with multilobed or multiple nuclei (Fig. 4a, H/E). In addition, spindles surrounding metaphase chromosomes were abnormal, displaying multiple spindle poles and conflicting attachments to condensed chromosomes (Fig. 4a, DNA/ α -tub). Fluorescent *in situ* hybridization (FISH) enumeration of chromosome X and 17 centromeres showed that the enlarged cells were hyperdiploid, that is, having more than the normal number of chromosomes X and 17 (Fig. 4a, X/17). Therefore, we

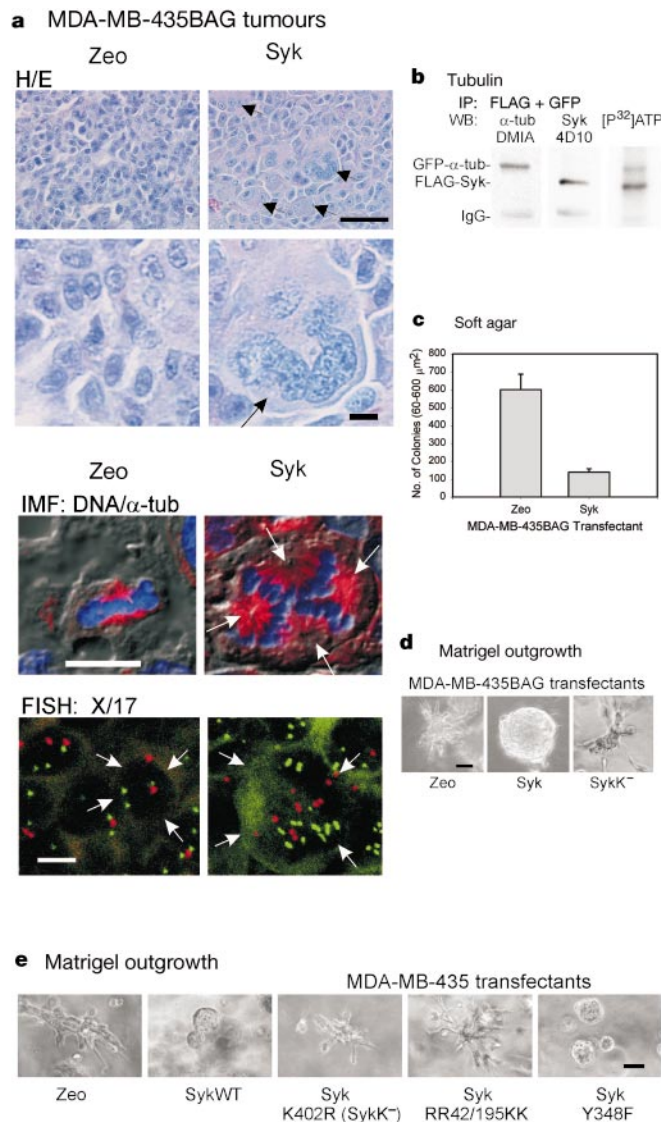


Figure 4 Biological consequences of Syk transfection in MDA-MB-435 cells. **a**, Paraffin sections of experimentally induced tumours stained with haematoxylin/eosin (H/E), anti- α -tubulin (red) and ToPro3 (DNA, blue) and FISH for chromosomes 17 (red) and X (green). Arrows indicate enlarged cells (in H/E), spindle poles (in IMF) or margins of individual cells (in FISH). Scale bars are 5 μm except in low-magnification H/E stains (20 μm).

b, Phosphorylation of α -tubulin by Syk in an immunocomplex kinase assay with epitope-tagged proteins in Cos1 cells. **c**, Anchorage-independent cell growth (mean $\pm$ s.e.). **d**, **e**, Outgrowth in Matrigel of cells transfected with wild-type or mutated Syk. Scale bar, 100 μm .

propose that tumour growth is blocked after abnormal mitosis and failure of subsequent cytokinesis in cells transfected with Syk. We showed that α -tubulin is phosphorylated by Syk in an immunocomplex kinase assay *in vitro* (Fig. 4b). α -Tubulin is also an *in vivo* substrate for Syk in activated B cells¹⁵. Moreover, it is phosphorylated on a tyrosine residue in the carboxy-terminal region containing the sites of interaction with the microtubule-associated proteins that regulate microtubule assembly and function¹⁵. Syk might therefore alter the microtubule/tubulin monomer equilibrium and ultimately affect mitosis. In agreement with our observations, abnormal mitoses due to altered microtubule organization have been observed in breast tumours¹⁶.

In vitro, all MDA-MB-435BAG transfectants showed an indistinguishable logarithmic proliferation on plastic substratum (data not shown); however, the MDA-MB-435BAG/Syk cells displayed a significantly decreased ability to grow in anchorage-independent conditions as compared to the control (Zeo) transfectants ($P < 0.005$, Student's *t*-test) (Fig. 4c). We observed that the ability of breast cancer cell lines to form invasive colonies in Matrigel, a naturally occurring basement membrane matrix, was inversely correlated with their Syk expression (Table 1; ref. 8). Correspondingly, all MDA-MB-435BAG/Zeo colonies displayed an invasive morphology with single cells penetrating the surrounding Matrigel matrix (Fig. 4d). The Syk-expressing colonies, in contrast, exhibited a well-delineated, non-invasive spherical appearance. Stable expression of a kinase-negative Syk (Syk^K) by these cells was not sufficient to suppress invasive outgrowth.

In summary, Syk expression specifically inhibits anchorage-independent proliferation and abates extracellular matrix invasion *in vitro*. As biological effects with Syk-transfected cells were only observed in animal experiments and in Matrigel but not on plastic substratum, we propose that the extracellular matrix might contain the signals that activate Syk¹⁷. We then generated several FLAG-tagged Syk mutants and tested their effect on Matrigel outgrowth after stable transfection in the Syk-negative MDA-MB-435 cells (Fig. 4e). Mutant Syk proteins with functionally inactive SH2 domains (RR42/195KK)¹⁸ or lacking kinase activity (K402R) lost their capacity to suppress the invasive outgrowth of cells in Matrigel. However, it was ineffective to mutate an autophosphorylated tyrosine residue in the Syk linker region (Y348F) that is critical for its interaction with the Vav¹⁹ and the PLC- γ 1 SH2 domains²⁰. This suggests that other Syk effectors may be necessary for the suppression of invasive outgrowth in Matrigel.

In conclusion, we have shown that the Syk tyrosine kinase is commonly expressed by normal breast epithelial cells, that loss of Syk expression is associated with the acquisition of a malignant breast tumour phenotype, and that Syk may directly act as a tumour suppressor, presumably by controlling cell division. We therefore propose that loss of Syk expression may function with other genetic alterations in the multistep process of breast tumour development and progression. This negative correlation is unexpected when considered with the current concept that tyrosine kinases are positively associated with tumorigenesis (for example, c-erbB2). Interestingly, allelic loss on chromosome 9q22, the human Syk locus, has been reported to be associated with lymph node metastasis of primary breast cancer²¹. As Syk knockout mice die perinatally, the effects of loss of Syk expression on breast development have not been not observed^{22,23}, but knockout mice lacking c-Cbl, a regulator of Syk, develop mammary epithelial hyperplasia²⁴. □

Methods

Cells lines and tissue samples

The MCF10A and K562 cell lines were obtained from the ATCC (Manassas, VA). All other human breast cancer cell lines were provided by the Tissue Culture Shared Resources of the Lombardi Cancer Center. Human tissue samples were obtained from the Cooperative Human Tissue Network (Univ. Pennsylvania, Philadelphia), the Lombardi Cancer Center Histopathology and Tissue shared resource, and C. Theillet (CNRS UMR 5535, CRLC Val d'Aurelle, Montpellier, France). Human mammary gland total RNA was purchased from Clontech.

Immunoprecipitation, western blotting and *in vitro* kinase activity

Cells were lysed as described⁶. Syk was immunoprecipitated and detected on nitrocellulose membrane by western blotting with commercial antibodies (Santa Cruz Biotechnology) and enhanced chemiluminescence (Amersham Pharmacia). The following antibodies were used: mouse monoclonal anti-phosphotyrosine (4G10; UBI, Lake Placid, NY), anti- α -tubulin (DM1A; Neomarkers, Fremont, CA), anti-FLAG antibody (M2; Sigma) and the rabbit polyclonal anti-GFP (Torrey Pines Biolabs, San Diego, CA). The autophosphorylation capacity of immunoprecipitated Syk was evaluated by the *in vitro* incorporation of 10 μ Ci [γ -³²P]ATP. Pervanadate (1 mM Na₂VO₄ + 1 mM H₂O₂) treatment of cells was carried out for 15 min at 37 °C.

PCR and Southern hybridization

The following human Syk primers were used for RT-PCR and genomic PCR: Syk-S1 forward, 5'-CATGTCAAGGATAAGAATCATCATAGA-3'; Syk-S3 forward, 5'-GTGTTTGCTA GTTACCAACATTACGC-3'; Syk-S5 forward, 5'-AAAGAAGTTCGACACGCTCTGG-3'; and Syk-AS1 reverse, 5'-AGTTCACCACTCATAGTAGTAATT-3'. Primers amplifying specifically porcine Syk were Syk-S4 forward, 5'-TGGGCAACGGGAATGATA-3'; Syk-S6 forward, 5'-TGTCAGGAGAAGGCGCAGG-3'; and Syk-AS4 reverse, 5'-ATTGGCCTCATTTTCAGGATCC-3'. Human β -actin primers correspond to nucleotides 936–955 (forward) and 1170–1189 (reverse). Human GAPDH primers correspond to nucleotides 67–86 (forward), 504–523 (reverse) and 361–380 (internal oligonucleotide). Southern hybridization was performed on Nylon membrane and Syk was detected with a biotinylated internal oligonucleotide (Syk-Biot., 5'-AGTTGATG-CATTCCG GAGCG-3'). For Southern hybridization of LCM/RT-PCR samples, internal primers were radiolabelled with [³²P]dATP using T4 polynucleotide kinase.

In situ hybridization

A 140-bp nucleotide fragment of the human Syk cDNA (nucleotides 472–611) was subcloned into the pGEM9Zf vector (Promega) and used to generate digoxigenin-UTP labelled cRNA probes by *in vitro* transcription. *In situ* hybridization was performed as described²⁵ and the signal revealed with alkaline phosphatase and NBT/BCIP substrate (Boehringer). A *lacZ* non-sense probe was used as negative control.

Laser capture microdissection (LCM)

Laser microdissection of a 30 μ m diameter capture area was done as described²⁶ using a Pixcell LCM system (Arcturus Engineering, Mountain View, California). Total RNA was retrieved and used for RT-PCR.

Vector construction and mutagenesis

The wild-type human Syk cDNA and the mutated kinase-deficient porcine Syk cDNA (K395R) were respectively provided by S. Yagi (Tonen Corp. Res. Dev. Lab., Saitama, Japan)²⁷ and T. Kurosaki (Kansai Medical Univ., Moriguchi, Japan)¹⁴. Both cDNAs were recombined into the pcDNA3.1-Zeo (+) mammalian expression vector (Invitrogen), verified by sequencing and transfected using electroporation. After selection in Zeocin (800 μ g ml⁻¹), 10–50 stably transfected clones were obtained and pooled to avoid unrepresentative subcloning artefacts. Alternatively, the human Syk coding sequence was amplified by PCR and cloned into the pCAF1 expression vector (gift of P. Burbelo, Georgetown Univ. Medical Center, Washington DC) to encode an amino-terminal FLAG-tagged Syk. Point mutations in the Syk coding sequence were generated (QuickChange, Stratagene) and verified by sequencing. Mutant Syk cDNAs were stably transfected using SuperFect reagent (Qiagen) and selected in G418 (800 μ g ml⁻¹) and their expression and catalytic activity were verified. The pEGFP- α -tubulin vector was purchased from Clontech.

Animal experiments

To determine their tumorigenicity, cells were injected subcutaneously (two to five million cells) into the mammary fat pad of 4–6-week-old female NCR *nu/nu* mice that were intact or ovariectomized and supplemented with a subcutaneously implanted 17- β -oestradiol pellet (Innovative Research, Rockville, Maryland)⁸. Primary tumours and organs (lungs, liver, spleen, lymph nodes) were fixed or snap frozen for immunohistochemistry or whole organ X-gal staining¹². To determine their metastatic capacity, two million cells were injected into the tail vein. After 4 weeks, the number of lung colonies were counted using a dissecting microscope and histologically confirmed.

TUNEL, immunohistochemistry and FISH

Apoptosis was detected by TUNEL using the TACS 2 TdT kit (Trevigen, Gaithersburg, Maryland) and quantified using Optimas 5.2 image analysis software⁶ (Bothell, Washington). α -Tubulin was visualized using monoclonal mouse anti- α -tubulin antibodies (clone B-5-1-2; Sigma) followed by Texas Red anti-mouse (Jackson Laboratories) and counterstained for DNA using ToPro3 (Molecular Probes). FISH for chromosomes 17 and X was done using directly labelled centromeric probes (Vysis, Downers Grove, Illinois)²⁸.

In vitro assays

Anchorage-independent growth in soft agar was carried out in triplicate as described²⁹ and counted with an Omnicon 3800 tumour colony analyser (Imaging Products International, Chantilly, Virginia). Matrigel outgrowth was assessed as described⁸ and observed on a Zeiss IM35 inverted microscope.

Received 17 February; accepted 24 May 2000.

- Cheng, A. M. & Chan, A. C. Protein tyrosine kinases in thymocyte development. *Curr. Opin. Immunol.* **9**, 528–533 (1997).
- Kurosaki, T. Molecular mechanisms in B cell antigen receptor signaling. *Curr. Opin. Immunol.* **9**, 309–318 (1997).
- Chu, D. H., Morita, C. T. & Weiss, A. The Syk family of protein tyrosine kinases in T-cell activation and development. *Immunol. Rev.* **165**, 167–180 (1998).
- Indik, Z. K., Park, J. G., Pan, X. Q. & Schreiber, A. D. Induction of phagocytosis by a protein tyrosine kinase. *Blood* **85**, 1175–1180 (1995).
- Fluck, M., Zurcher, G., Andres, A. C. & Ziemiecki, A. Molecular characterization of the murine Syk protein tyrosine kinase cDNA, transcripts and protein. *Biochem. Biophys. Res. Commun.* **213**, 273–281 (1995).
- Coopman, P. J., Do, M. T. H., Thompson, E. W. & Mueller, S. C. Phagocytosis of cross-linked gelatin matrix by human breast carcinoma cells correlates with their invasive capacity. *Clin. Canc. Res.* **4**, 507–515 (1998).
- Coopman, P. J., Thomas, D. M., Gehlsen, K. R. & Mueller, S. C. Integrin $\alpha 3 \beta 1$ participates in the phagocytosis of extracellular matrix molecules by human breast cancer cells. *Mol. Biol. Cell* **7**, 1789–1804 (1996).
- Thompson, E. W. *et al.* Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. *J. Cell Physiol.* **150**, 534–544 (1992).
- Law, C. L. *et al.* Molecular cloning of human Syk. A B cell protein-tyrosine kinase associated with the surface immunoglobulin M-B cell receptor complex. *J. Biol. Chem.* **269**, 12310–12319 (1994).
- Nagai, K., Inazu, T. & Yamamura, H. p72syk is activated by vanadate plus H_2O_2 in porcine platelets and phosphorylates GTPase activating protein on tyrosine residue(s). *J. Biochem. (Tokyo)* **116**, 1176–1181 (1994).
- Fagnoli, J. *et al.* Syk mutation in Jurkat E6-derived clones results in lack of p72^{syk} expression. *J. Biol. Chem.* **270**, 26533–26537 (1995).
- Brunner, N. *et al.* lacZ transduced human breast cancer xenografts as an *in vivo* model for the study of invasion and metastasis. *Eur. J. Cancer* **28A**, 1989–1995 (1992).
- Kurebayashi, J. *et al.* Quantitative demonstration of spontaneous metastasis by MCF-7 human breast cancer cells cotransfected with fibroblast growth factor 4 and lacZ. *Cancer Res.* **53**, 2178–2187 (1993).
- Takata, M. *et al.* Tyrosine kinases Lyn and Syk regulate B cell receptor-coupled Ca^{2+} mobilization through distinct pathways. *EMBO J.* **13**, 1341–1349 (1994).
- Peters, J. D., Furlong, M. T., Asai, D. J., Harrison, M. L. & Geahlen, R. L. Syk, activated by cross-linking the B-cell antigen receptor, localizes to the cytosol where it interacts with and phosphorylates alpha-tubulin on tyrosine. *J. Biol. Chem.* **271**, 4755–4762 (1996).
- Lingle, W. L. & Salisbury, J. L. Altered centrosome structure is associated with abnormal mitoses in human breast tumors. *Am. J. Pathol.* **155**, 1941–1951 (1999).
- Miranti, C. K., Leng, L., Maschberger, P., Brugge, J. S. & Shattil, S. J. Identification of a novel integrin signaling pathway involving the kinase Syk and the guanine nucleotide exchange factor Vav1. *Curr. Biol.* **8**, 1289–1299 (1998).
- Williams, S. *et al.* Reconstitution of T cell antigen receptor-induced Erk2 kinase activation in Lck-negative JCaM1 cells by Syk. *Eur. J. Biochem.* **245**, 84–90 (1997).
- Deckert, M., Tartare-Deckert, S., Couture, C., Mustelin, T. & Altman, A. Functional and physical interactions of Syk family kinases with the Vav proto-oncogene product. *Immunity* **5**, 591–604 (1996).
- Law, C. L., Chandran, K. A., Sidorenko, S. P. & Clark, E. A. Phospholipase C-gamma1 interacts with conserved phosphotyrosyl residues in the linker region of Syk and is a substrate for Syk. *Mol. Cell. Biol.* **16**, 1305–1315 (1996).
- Minobe, K. *et al.* Allelic loss on chromosome 9q is associated with lymph node metastasis of primary breast cancer. *Jpn. J. Cancer Res.* **89**, 916–922 (1998).
- Turner, M. *et al.* Perinatal lethality and blocked B-cell development in mice lacking the tyrosine kinase Syk. *Nature* **378**, 298–302 (1995).
- Cheng, A. M. *et al.* Syk tyrosine kinase required for mouse viability and B-cell development. *Nature* **378**, 303–306 (1995).
- Murphy, M. A. *et al.* Tissue hyperplasia and enhanced T-cell signalling via ZAP-70 in c-Cbl-deficient mice. *Mol. Cell Biol.* **18**, 4872–4882 (1998).
- Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
- Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
- Yagi, S., Suzuki, K., Hasegawa, A., Okumura, K. & Ra, C. Cloning of the cDNA for the deleted Syk kinase homologous to Zap-70 from human basophilic leukemia cell line (KU812). *Biochem. Biophys. Res. Commun.* **200**, 28–34 (1994).
- Pinkel, D. *et al.* Cytogenetic analysis by *in situ* hybridization with fluorescently labeled nucleic acid probes. *Cold Spring Harbor Symp. Quant. Biol.* **51**, 151–157 (1986).
- Fang, W., Hartmann, N., Chow, D. T., Riegel, A. T. & Wellstein, A. Pleiotrophin stimulates fibroblasts and endothelial and epithelial cells and is expressed in human cancer. *J. Biol. Chem.* **267**, 25889–25897 (1992).

Acknowledgements

We gratefully acknowledge S. Yagi and T. Kurosaki who have provided the Syk cDNAs. We thank P. Burbelo for the pCAF1 vector; F. Kern for ML20 cells; J. Zwiebel for MDA-MB-435BAG cells; C. Theillet for tumour samples; A. Wright, F. Onajafe and M. Dai for technical assistance; S. Artero for statistics; and B. Singh and P. Roger for pathology advice. We also thank J. Brugge, N. Taylor, A. Wellstein and M. Lippman for comments on the manuscript. This work was supported in part by NIH grants (to S.C.M. and S.W.M.) and, in part, by the Lombardi Cancer Center shared resources supported by a US Public Health Service Grant.

Correspondence and requests for materials should be addressed to S.C.M. (e-mail: muellers@gunet.georgetown.edu).

Molecular portraits of human breast tumours

Charles M. Perou^{*†}, Therese Sørlie^{†‡}, Michael B. Eisen^{*}, Matt van de Rijn[§], Stefanie S. Jeffrey^{||}, Christian A. Rees^{*}, Jonathan R. Pollack[¶], Douglas T. Ross[¶], Hilde Johnsen[‡], Lars A. Akslen[#], Øystein Fluge[☆], Alexander Pergamenschikov^{*}, Cheryl Williams^{*}, Shirley X. Zhu[§], Per E. Lønning^{**}, Anne-Lise Børresen-Dale[‡], Patrick O. Brown^{††} & David Botstein^{*}

^{*} Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

[‡] Department of Genetics, The Norwegian Radium Hospital, N-0310 Montebello Oslo, Norway

[§] Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA

^{||} Department of Surgery, Stanford University School of Medicine, Stanford, California 94305, USA

[¶] Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

[#] Department of Pathology, The Gade Institute, Haukeland University Hospital, N-5021 Bergen, Norway

[☆] Department of Molecular Biology, University of Bergen, N-5020 Bergen, Norway

^{**} Department of Oncology, Haukeland University Hospital, N-5021 Bergen, Norway

^{††} Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA

[†] These authors contributed equally to this work

Human breast tumours are diverse in their natural history and in their responsiveness to treatments¹. Variation in transcriptional programs accounts for much of the biological diversity of human cells and tumours. In each cell, signal transduction and regulatory systems transduce information from the cell's identity to its environmental status, thereby controlling the level of expression of every gene in the genome. Here we have characterized variation in gene expression patterns in a set of 65 surgical specimens of human breast tumours from 42 different individuals, using complementary DNA microarrays representing 8,102 human genes. These patterns provided a distinctive molecular portrait of each tumour. Twenty of the tumours were sampled twice, before and after a 16-week course of doxorubicin chemotherapy, and two tumours were paired with a lymph node metastasis from the same patient. Gene expression patterns in two tumour samples from the same individual were almost always more similar to each other than either was to any other sample. Sets of co-expressed genes were identified for which variation in messenger RNA levels could be related to specific features of physiological variation. The tumours could be classified into subtypes distinguished by pervasive differences in their gene expression patterns.

We proposed that the phenotypic diversity of breast tumours might be accompanied by a corresponding diversity in gene expression patterns that we could capture using cDNA microarrays. Systematic investigation of gene expression patterns in human breast tumours might then provide the basis for an improved molecular taxonomy of breast cancers. We analysed gene expression patterns in grossly dissected normal or malignant human breast tissues from 42 individuals (36 infiltrating ductal carcinomas, 2 lobular carcinomas, 1 ductal carcinoma *in situ*, 1 fibroadenoma and 3 normal breast samples). Fluorescently labelled (Cy5) cDNA was prepared from mRNA from each experimental sample. We prepared cDNA, labelled using a second distinguishable fluorescent nucleotide (Cy3), from a pool of mRNAs isolated from 11 different

Twenty of the forty breast tumours examined were sampled twice,

resection of the remaining tumour. In addition, primary tumours



Figure 1 Variation in expression of 1,753 genes in 84 experimental samples. Data are presented in a matrix format: each row represents a single gene, and each column an experimental sample. In each sample, the ratio of the abundance of transcripts of each gene to the median abundance of the gene's transcript among all the cell lines (left panel), or to its median abundance across all tissue samples (right panel), is represented by the colour of the corresponding cell in the matrix. Green squares, transcript levels below the median; black squares, transcript levels equal to the median; red squares, transcript levels greater than the median; grey squares, technically inadequate or missing data. Colour saturation reflects the magnitude of the ratio relative to the median for each set of samples (see scale, bottom left; and Supplementary Information Fig. 4). **a**, Dendrogram representing similarities in the expression patterns between experimental samples. All 'before and after' chemotherapy pairs that were clustered on terminal branches are highlighted in red; the two primary tumour/lymph node metastasis pairs in light blue; the three clustered normal breast samples in light green. Branches representing the four breast luminal epithelial cell lines are shown in dark blue; breast basal epithelial cell lines in orange, the endothelial cell lines in dark yellow, the mesenchymal-like cell lines in dark green, and the lymphocyte-derived cell lines in brown. **b**, Scaled-down representation of the 1,753-gene cluster diagram; coloured bars to the right identify the locations of the inserts displayed in **c–j**. **c**, Endothelial cell gene expression cluster; **d**, stromal/fibroblast cluster; **e**, breast basal epithelial cluster; **f**, B-cell cluster; **g**, adipose-enriched/normal breast; **h**, macrophage; **i**, T-cell; **j**, breast luminal epithelial cell.

from two patients were also paired with a lymph node metastasis from the same patient. To help interpret the variation in expression patterns seen in the tumour samples, we also characterized 17 cultured cell lines (with one cell line cultured under three different conditions), which provided models for many of the cell types encountered in these tissue samples. In total, we analysed 84 cDNA microarray experiments (see Supplementary Information, Table 2; the primary data tables can be obtained at <http://genome-www.stanford.edu/molecularportraits/>).

A hierarchical clustering method was used to group genes on the basis of similarity in the pattern with which their expression varied over all samples⁵. The same clustering method was used to group the experimental samples (cell lines and tissues separately) on the basis of similarity in their patterns of expression. We focus first on a set of 1,753 genes (about 22% of the 8,102 genes analysed), whose transcripts varied in abundance by at least fourfold from their median abundance in this sample set in at least three of the samples (Fig. 1; see Supplementary Information Fig. 4 for the complete cluster diagram).

Three striking features of the gene expression patterns of these tumours are evident in Fig. 1. First, the tumours show great variation in their patterns of gene expression. Second, this variation

is multidimensional; that is, many different sets of genes show mainly independent patterns of variation. Third, these patterns have a pervasive order reflecting relationships among the genes, relationships among the tumours and connections between specific genes and specific tumours.

The hierarchical clustering algorithm organizes the experimental samples only on the basis of overall similarity in their gene expression patterns; these relationships are summarized in a dendrogram (Fig. 1a), in which the pattern and length of the branches reflects the relatedness of the samples⁵. Fifteen of the twenty before and after doxorubicin pairs (red dendrogram branches), and both primary tumour/lymph node metastasis pairs (light blue branches) were clustered together on terminal branches in the dendrogram; that is, despite an interval of 16 weeks, independent surgical procedures and cytotoxic chemotherapy, independent samples taken from the same tumour were in most cases recognizably more similar to each other than either was to any of the other samples. In three instances (Norway 47, 61 and 101), the 'after' chemotherapy specimens clustered in a branch of the dendrogram that also contained the three normal breast samples; we know from the clinical data that these tumours were 3 of the 20 tumours that were classified as doxorubicin 'responders' (data not shown). An analysis of the relationship between gene expression and correlations with clinical data will be reported elsewhere (T.S. *et al.*, manuscript in preparation).

The 'molecular portraits' revealed in the patterns of gene expression not only uncovered similarities and differences among the tumours, but in many cases pointed to a biological interpretation. Variation in growth rate, in the activity of specific signalling pathways, and in the cellular composition of the tumours were all reflected in the corresponding variation in the expression of specific subsets of genes. The largest distinct cluster of genes within the 1,753-gene cluster diagram was the 'proliferation cluster' (Supplementary Information Fig. 5), which is a group of genes whose levels of expression correlate with cellular proliferation rates^{3,6}. Expression of this cluster of genes varied widely among the tumour samples, and was generally well correlated with the mitotic index. As one might expect, this cluster also included the genes encoding two widely used immunohistochemical markers of cell proliferation (Ki-67 and PCNA).

Several groups of co-expressed genes provided views of the activities of specific signalling and/or regulatory systems. A large cluster of genes regulated by the interferon pathway (including *STAT1*) showed substantial variation in expression among the tumours, as was previously observed in a smaller set of breast

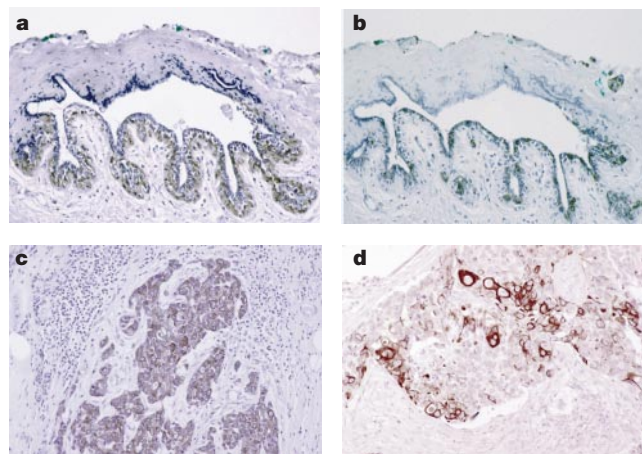


Figure 2 Breast tissue immunohistochemistry. **a**, Normal mammary duct using antibodies against the basal keratins 5/6. **b**, Normal mammary duct using antibodies against the luminal keratins 8/18 (adjacent tissues sections were used in **a** and **b**). **c**, Tumour

Stanford 16 using antibodies against keratins 8/18. **d**, Tumour New York 3 using antibodies against keratins 5/6.

tumours⁶. Variation in expression of the oestrogen receptor- α gene (ER) correlated well with the direct clinical measurement of the ER protein levels in the tumours (Supplementary Information Table 3; concordance in 36/38 samples), and paralleled variation in the expression of a larger group of genes that included three other

transcription factors (GATA-binding protein 3 (refs 7, 8), X-box binding protein 1 and hepatocyte nuclear factor 3 α). *HER2/neu*, also known as *Erb-B2*, is overexpressed in 20–30% of all breast tumours, usually associated with DNA amplification of the *Erb-B2* locus^{9,10}. Notably, most of the other genes contained within the

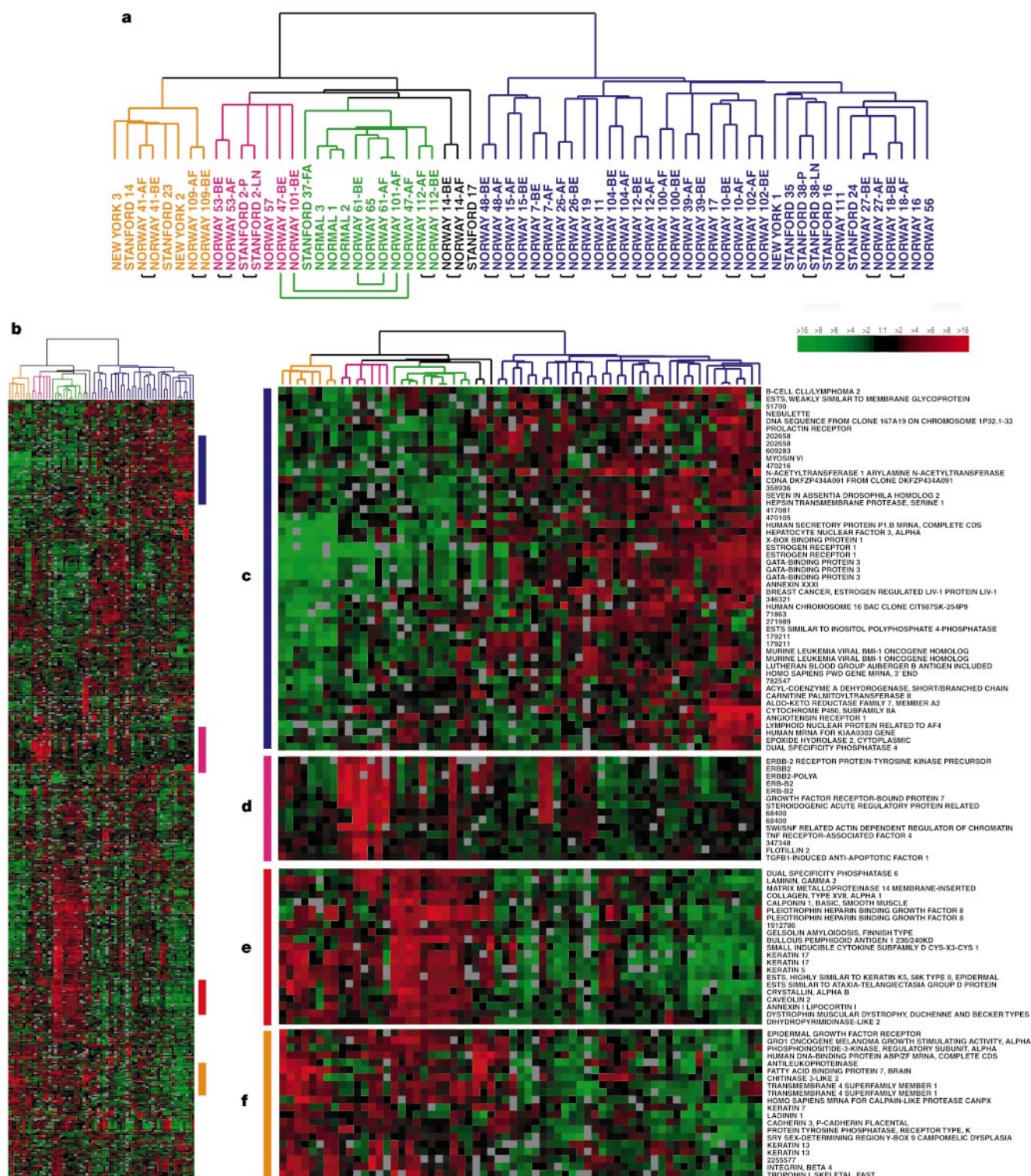


Figure 3 Cluster analysis using the 'intrinsic' gene subset. Two large branches were apparent in the dendrogram, and within these large branches were smaller branches for which common biological themes could be inferred. Branches are coloured accordingly: basal-like, orange; *Erb-B2*+, pink; normal-breast-like, light green; and luminal epithelial/ER+, dark blue. **a**, Experimental sample associated cluster dendrogram. Small black bars beneath the dendrogram identify the 17 pairs that were matched by this hierarchical clustering; larger green bars identify the positions of the three pairs that were not matched by the clustering. **b**, Scaled-down representation of the intrinsic cluster diagram (see Supplementary Information Fig. 6). **c**, Luminal epithelial/ER gene cluster. **d**, *Erb-B2* overexpression cluster. **e**, Basal epithelial cell associated cluster containing keratins 5 and 17. **f**, A second basal epithelial-cell-enriched gene cluster.

Erb-B2 cluster were located in this same region of chromosome 17, and were also amplified on the genomic DNA level (ref. 10; and J.R.P., unpublished data). Finally, a cluster of genes that included *c-Fos* and *JunB* co-varied in expression among the tumour specimens. We have found that this subset of genes is characteristically induced by prolonged handling of the samples after surgical resection (M.v.d.R. and C.M.P., unpublished data).

Human breast tumours are histologically complex tissues, containing a variety of cell types in addition to the carcinoma cells¹¹. In analysing the gene expression patterns in solid human tumours, we used two lines of reasoning to infer the lineage of the cells that accounted for the apparently cell-type-specific expression of particular clustered groups of genes. First, such clusters included genes whose expression patterns have been previously characterized and that consistently pointed to a specific cell type. Second, these inferences were often corroborated by comparable expression of the same cluster in one or more of the cultured cell lines. Thus, eight independent clusters of genes appeared to reflect variation in specific cell types present within the tumours (Fig. 1c–j).

(1) Endothelial cells: a cluster of genes characteristically expressed by endothelial cells, including CD34, CD31 and von Willebrand factor were also strongly expressed in the two endothelial cell lines HUVEC and HMVEC (Fig. 1c). (2) Stromal cells: a previously characterized cluster of genes that included several isoforms of collagen showed significant variation in expression among samples (Fig. 1d)^{3,6}. (3) Adipose-enriched/normal breast cells: a cluster of genes including fatty-acid binding protein 4 and PPAR γ may represent the presence of adipose cells (Fig. 1g). (4) B lymphocytes: variation in expression of a cluster of genes that were highly expressed in the multiple myeloma-derived cell line RPMI-8226, including many immunoglobulin genes, appears to represent variable B-cell infiltration (Fig. 1f). (5) T lymphocytes: a cluster of genes including CD3 δ and two subunits of the T-cell receptor were highly expressed in the T-cell leukaemia-derived cell line MOLT-4 and probably indicate T-cell infiltrates (Fig. 1i). (6) Macrophages: a cluster of genes that appeared to be markers of macrophage/monocytes included CD68, acid phosphatase 5, chitinase and lysozyme (Fig. 1h).

Two distinct types of epithelial cell are found in the human mammary gland: basal (and/or myoepithelial) cells and luminal epithelial cells^{11,12}. These two cell types are conveniently distinguished immunohistochemically; basal epithelial cells can be stained with antibodies to keratin 5/6 (Fig. 2a), whereas luminal epithelial cells stain with antibodies against keratins 8/18 (Fig. 2b). Many genes were expressed by one of these two cell lineages, but not by the other (Fig. 1e and j). The gene expression cluster characteristic of basal epithelial cells included keratin 5, keratin 17, integrin- β 4 and laminin (Fig. 1e)¹¹. The gene expression cluster characteristic of the luminal cells was anchored by the previously noted cluster of transcription factors that included ER (Fig. 1j).

One goal of this study was to develop a system for classifying tumours on the basis of their gene expression patterns. The subset of genes shown in Fig. 1 was not necessarily optimal for this purpose, as the choice of genes whose expression levels provided the basis for the ordering of the tumour samples determined which phenotypic relationships among the tumours were reflected in the clustering patterns. We therefore selected an alternative subset of genes to use as the basis for a new clustering analysis.

The rationale behind this alternative gene subset was that specific features of a gene expression pattern that are to be used to classify tumours should be similar in any sample taken from the same tumour, and they should vary among different tumours. The 22 paired samples provided a unique opportunity for a deliberate and systematic search for such genes. From the genes whose expression was well measured in the 65 tissue samples, we selected a subset of 496 genes (termed the 'intrinsic' gene subset) that consisted of genes with significantly greater variation in expression between different

tumours than between paired samples from the same tumour (see Supplementary Information). When variation in expression of this set of genes was used to order the tissue samples (Fig. 3; and Supplementary Information Fig. 6), 17 of the 20 'before and after' doxorubicin pairs were grouped together as were both of the tumour/lymph node metastasis pairs. Qualitatively similar sample clustering patterns were obtained when a second gene subset that focused on genes expressed by epithelial cell types, and which had only 25% overlap with the intrinsic gene subset, was used (data not shown).

The division of the tissue samples into two subgroups was a striking feature of the intrinsic gene subset cluster analysis (Fig. 3a). As a test of the robustness of this division, we applied the 'weighted voting' method¹³. This algorithm recapitulated the sorting of the tissue samples between these two subgroups for all but 1 of the 65 samples (data not shown). It is important to note, however, that there is extensive residual variation in expression patterns within each of these two broad subgroups. Indeed, many of the finer subdivisions probably have important biological properties (see below).

The two dendrogram branches in Fig. 3 largely separate the tumour samples into those that were clinically described as ER positive (blue) and those that were ER negative (other colours). The tumours in the ER+ group were characterized by the relatively high expression of many genes expressed by breast luminal cells (Fig. 3c). This connection was further corroborated using immunohistochemical analysis and antibodies against the luminal cell keratins 8/18 (Fig. 2c). With one exception, none of the tumours in this group expressed *Erb-B2* at high levels (Fig. 3d).

Many of the genes characteristic of breast basal epithelial cells were also highly expressed in a group of six clustered tumours (Fig. 3e). To corroborate the 'basal-like' characteristics of these tumours, we carried out immunohistochemistry using antibodies against the breast basal cell keratins 5/6 and 17. All six of these tumours showed staining for either keratins 5/6 or 17 or both (Fig. 2d). Notably, these six tumours also failed to express ER and most of the other genes that were usually co-expressed with it (Fig. 3c). Breast tumours that stain positive for basal keratins have been described^{14–16}, and basal keratins may account for 3–15% of all breast tumours^{15,17–19}; in this study, the incidence was 15% (6/40).

As mentioned above, overexpression of the *Erb-B2* oncogene was associated with the high expression of a specific subset of genes. We identified a cluster of tumours that was partially characterized by the high level of expression of this subset of genes (Fig. 3d). These tumours also showed low levels of expression of ER^{20,21} and of almost all of the other genes associated with ER expression—a trait they share with the basal-like tumours.

Several tumour samples and the single fibroadenoma tested (Fig. 3, light green), were clustered with a group of samples that also contained the three normal breast specimens (Fig. 3a). The 'normal breast' gene expression pattern is typified by the high expression of genes characteristic of basal epithelial cells and adipose cells, and the low expression of genes characteristic of luminal epithelial cells.

The number of clearly different molecular phenotypes observed among the breast tumours suggests that we are far from having a complete picture of the diversity of breast tumours. When hundreds (instead of tens) of breast tumours have been characterized, a more defined tumour classification is likely, and statistically significant relationships with clinical parameters should be uncovered. We were, however, able to identify four groups of samples that might be related to different molecular features of mammary epithelial biology (that is, ER+/luminal-like, basal-like, *Erb-B2*+ and normal breast). An important implication of this study is that the clinical designation of 'oestrogen receptor negative' breast carcinoma encompasses at least two biologically distinct subtypes of tumours (basal-like and *ErB-B2* positive), which may need to be treated as distinct diseases.

A striking conclusion from these data concerns the stability,

homogeneity and uniqueness of the 'molecular portraits' provided by the quantitative analysis of gene expression patterns. We infer that these portraits faithfully represent the 'tumour' itself, and not merely the particular tumour 'sample', because we could recognize the distinctive expression pattern of a tumour in independent samples. The finding that a metastasis and primary tumour were as similar in their overall pattern of gene expression as were repeated samplings of the same primary tumour, suggests that the molecular program of a primary tumour may generally be retained in its metastases. Finally, we have explicitly discussed only a tiny fraction of the genes whose expression patterns varied among these tumours. Attention to the thousands of individual genes that define the molecular portraits of each tumour, and learning to interpret their patterns of variation, will undoubtedly lead to a deeper and more complete understanding of breast cancers. □

Methods

Most of the techniques used in this work have been described elsewhere^{2,3,22,23}, and detailed protocols are available at (<http://cmgm.stanford.edu/pbrown/>). The methods and protocols are also included in the Supplementary Information, and the primary data tables can be obtained at (<http://genome-www.stanford.edu/molecularportraits/>).

Received 7 February; accepted 25 May 2000.

1. Tavassoli, F. A. & Schnitt, S. J. *Pathology of the Breast* (Elsevier, New York, 1992).
2. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303**, 179–205 (1999).
3. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24**, 227–235 (2000).
4. Aas, T. *et al.* Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nature Med.* **2**, 811–814 (1996).
5. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
6. Perou, C. M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**, 9212–9217 (1999).
7. Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O. & Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.* **27**, 1517–1523 (1999).
8. Hoch, R. V., Thompson, D. A., Baker, R. J. & Weigel, R. J. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer* **84**, 122–128 (1999).
9. Pauletti, G., Godolphin, W., Press, M. F. & Slamon, D. J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* **13**, 63–72 (1996).
10. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
11. Ronnov-Jessen, L., Petersen, O. W. & Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol. Rev.* **76**, 69–125 (1996).
12. Taylor-Papadimitriou, J. *et al.* Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. *J. Cell Sci.* **94**, 403–413 (1989).
13. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
14. Dairkee, S. H., Mayall, B. H., Smith, H. S. & Hackett, A. J. Monoclonal marker that predicts early recurrence of breast cancer. *Lancet* **1**, 514 (1987).
15. Dairkee, S. H., Puett, L. & Hackett, A. J. Expression of basal and luminal epithelium-specific keratins in normal, benign, and malignant breast tissue. *J. Natl Cancer Inst.* **80**, 691–695 (1988).
16. Malzahn, K., Mitze, M., Thoenes, M. & Moll, R. Biological and prognostic significance of stratified epithelial cytotokeratins in infiltrating ductal breast carcinomas. *Virchows Arch.* **433**, 119–129 (1998).
17. Guelstein, V. I. *et al.* Monoclonal antibody mapping of keratins 8 and 17 and of vimentin in normal human mammary gland, benign tumors, dysplasias and breast cancer. *Int. J. Cancer* **42**, 147–153 (1988).
18. Gusterson, B. A. *et al.* Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. *Cancer Res.* **42**, 4763–4770 (1982).
19. Nagle, R. B. *et al.* Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J. Histochem. Cytochem.* **34**, 869–881 (1986).
20. Berns, E. M. *et al.* Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumors: correlation with steroid receptors. *Eur. J. Cancer* **28**, 697–700 (1992).
21. Heintz, N. H., Leslie, K. O., Rogers, L. A. & Howard, P. L. Amplification of the c-erb B-2 oncogene and prognosis of breast adenocarcinoma. *Arch. Pathol. Lab. Med.* **114**, 160–163 (1990).
22. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
23. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank W. Gerald and L. Norton for the three New York tumour specimens; M. Stampfer and P. Yaswen for the 184 sample mRNAs; and members of the P. O. Brown, D. Botstein

and A.-L. Børresen-Dale labs for discussions. We are grateful to the NCI and the Howard Hughes Medical Institute who provided support for this research. C.M.P. is a SmithKline Beecham Pharmaceuticals Fellow of the Life Sciences Research Foundation. T.S. is a research fellow of the Norwegian Cancer Society. M.B.E. is an Alfred P. Sloan Foundation Postdoctoral Fellow in Computational Molecular Biology. D.T.R. is a Walter and Idun Berry Fellow. P.O.B. is an Associate Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.B.

(e-mail: botstein@genome.stanford.edu) or P.O.B.

(e-mail: pbrown@cmgm.stanford.edu).

Potential for biomolecular imaging with femtosecond X-ray pulses

Richard Neutze*, Remco Wouts*, David van der Spoel*, Edgar Weckert†‡ & Janos Hajdu*

* Department of Biochemistry, Biomedical Centre, Box 576, Uppsala University, S-75123 Uppsala, Sweden

† Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-76128, Germany

Sample damage by X-rays and other radiation limits the resolution of structural studies on non-repetitive and non-reproducible structures such as individual biomolecules or cells¹. Cooling can slow sample deterioration, but cannot eliminate damage-induced sample movement during the time needed for conventional measurements^{1,2}. Analyses of the dynamics of damage formation^{3–5} suggest that the conventional damage barrier (about 200 X-ray photons per Å² with X-rays of 12 keV energy or 1 Å wavelength²) may be extended at very high dose rates and very short exposure times. Here we have used computer simulations to investigate the structural information that can be recovered from the scattering of intense femtosecond X-ray pulses by single protein molecules and small assemblies. Estimations of radiation damage as a function of photon energy, pulse length, integrated pulse intensity and sample size show that experiments using very high X-ray dose rates and ultrashort exposures may provide useful structural information before radiation damage destroys the sample. We predict that such ultrashort, high-intensity X-ray pulses from free-electron lasers^{6,7} that are currently under development, in combination with container-free sample handling methods based on spraying techniques, will provide a new approach to structural determinations with X-rays.

Radiation damage is caused by X-ray photons depositing energy directly into the sample. At 1 Å wavelength, the photoelectric cross-section of carbon is about 10 times higher than its elastic-scattering cross-section, making the photoelectric effect the primary source of damage. The photoelectric effect is a resonance phenomenon in which a photon is absorbed and an electron ejected⁸, usually from a low-lying orbital of the atom (about 95% of the photoelectric events remove K-shell electrons from carbon, nitrogen, oxygen and sulphur), producing a hollow ion with an unstable electronic configuration. Relaxation is achieved through an electron from a higher shell falling into the vacant orbital. In heavy elements this usually gives rise to X-ray fluorescence, whereas in light elements the falling electron is more likely to give up its energy to another electron, which is then ejected in the Auger effect. Auger emission is predominant in carbon, nitrogen, oxygen and sulphur (> 95%)⁹; thus, most photoelectric events ultimately remove two electrons from these elements. These two electrons have different energies (~12 keV for photoelectrons and ~0.25 keV for Auger electrons),

‡ Present address: HASYLAB at DESY, Notkestr. 85, D-22607 Hamburg, Germany.

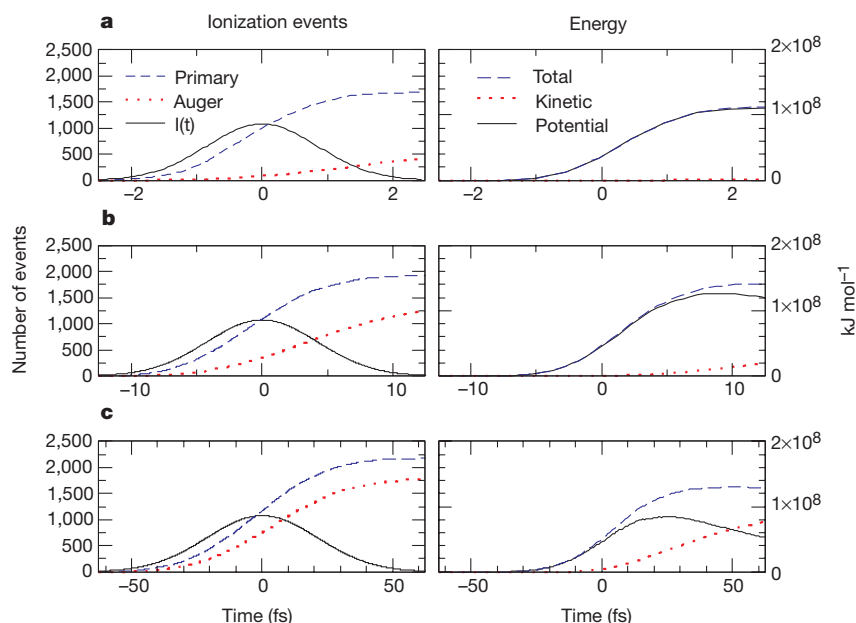


Figure 1 Ionization of a lysozyme molecule in intense X-ray pulses. The full width at half maximum (FWHM) of the pulse was 2 fs (**a**), 10 fs (**b**) or 50 fs (**c**). The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^6 photons per $\text{\AA}^2$) in all cases. The creation of a large number of positive charges owing to primary ionization by X-rays (photoelectric effect and Compton scattering), and the subsequent

Auger emissions (left) result in a rise in the potential (mainly electrostatic) energy of the sample (right). The degree of conversion of potential energy into kinetic energy during the X-ray exposure is inertia limited, and so depends strongly on the duration of the pulse. The gaussian shape of the X-ray pulse is indicated.

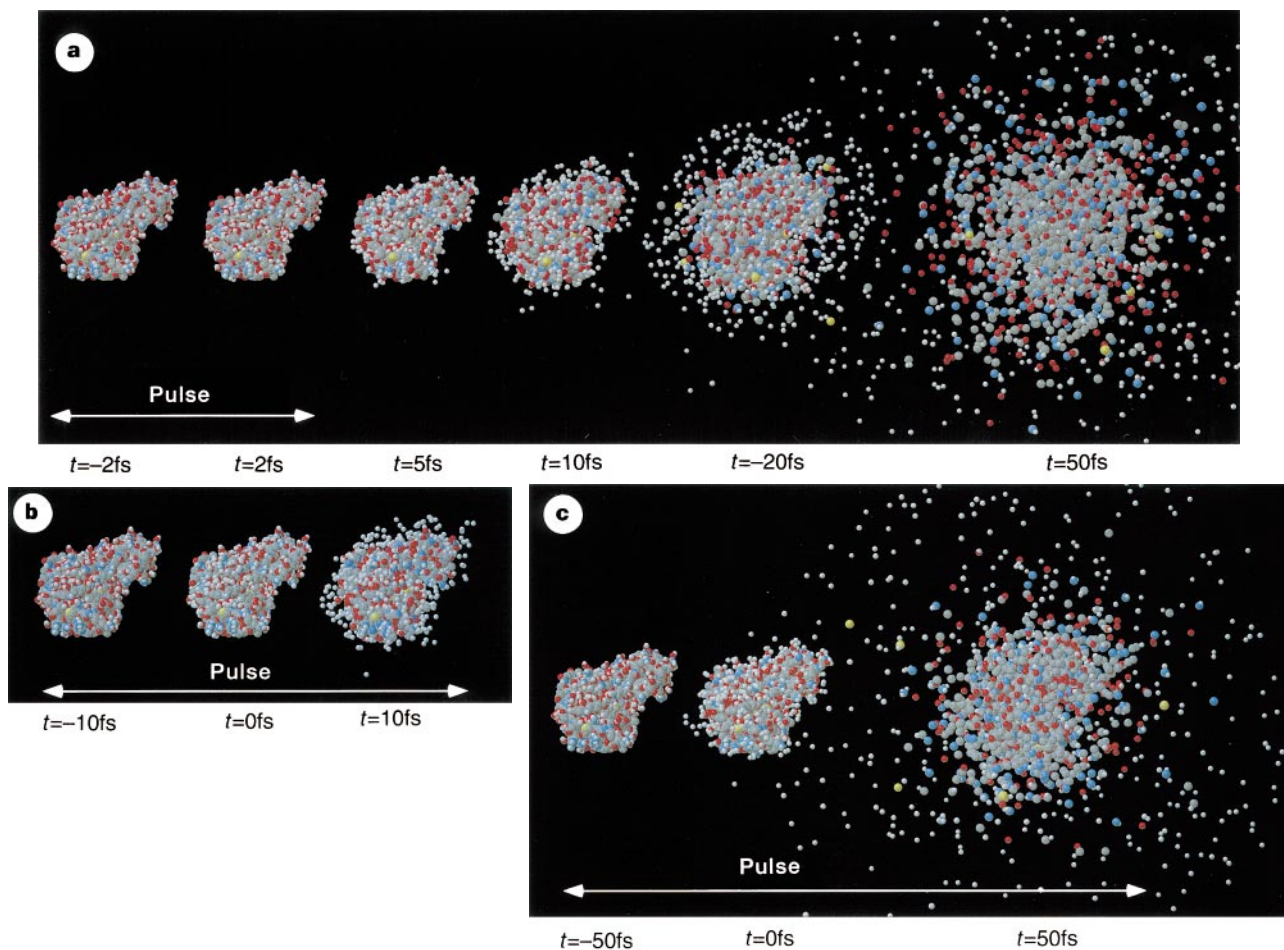


Figure 2 Explosion of T4 lysozyme (white, H; grey, C; blue, N; red, O; yellow, S) induced by radiation damage. The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^6 photons per $\text{\AA}^2$) in all cases. **a**, A protein exposed to an X-ray pulse with an FWHM of 2 fs, and disintegration followed in time. Atomic positions in the first two structures (before and after the pulse) are practically identical at this pulse length

because of an inertial delay in the explosion. $R_{\text{nuc}} = 3\%$, $R_{\text{elec}} = 11\%$ **b**, Lysozyme exposed to the same number of photons as in **a**, but the FWHM of the pulse was 10 fs. Images show the structure at the beginning, in the middle and near the end of the X-ray pulse. $R_{\text{nuc}} = 7\%$, $R_{\text{elec}} = 12\%$ **c**, Behaviour of the protein during an X-ray pulse with an FWHM of 50 fs. $R_{\text{nuc}} = 26\%$, $R_{\text{elec}} = 30\%$.

and are released at different times. Relevant K-hole lifetimes determined from Auger line-widths⁹ are 11.1 fs (C), 9.3 fs (N), 6.6 fs (O) and 1.3 fs (S). Shake-up excitations and interference between decay channels will modulate this picture.

Another effect is Compton (or inelastic) scattering, which represents a direct momentum transfer from an X-ray photon to a bound electron, so that the X-ray photon is scattered with a reduced energy. If the recoil energy taken up by the electron is greater than its shell binding energy, the atom will be ionized. The inelastic cross-section of carbon, nitrogen and oxygen is about 3% of the corresponding

photoelectric cross-sections^{10,11}, whereas the inelastic cross-section of hydrogen is much higher than its photoelectric cross-section.

The average velocities of photoelectrons (43 nm fs^{-1}) and Auger electrons (7 nm fs^{-1}) enable these electrons to escape the protein environment in less than a femtosecond during early phases of an exposure. At these velocities the inelastic electron scattering cross-sections for carbon¹ lie between 0.1 and $2.0 \text{ \AA}^2$, such that roughly one electron in five will deposit additional energy into the molecule from which it escapes, and may also remove outer-shell electrons¹². In late phases of an exposure, a significant fraction of the emitted

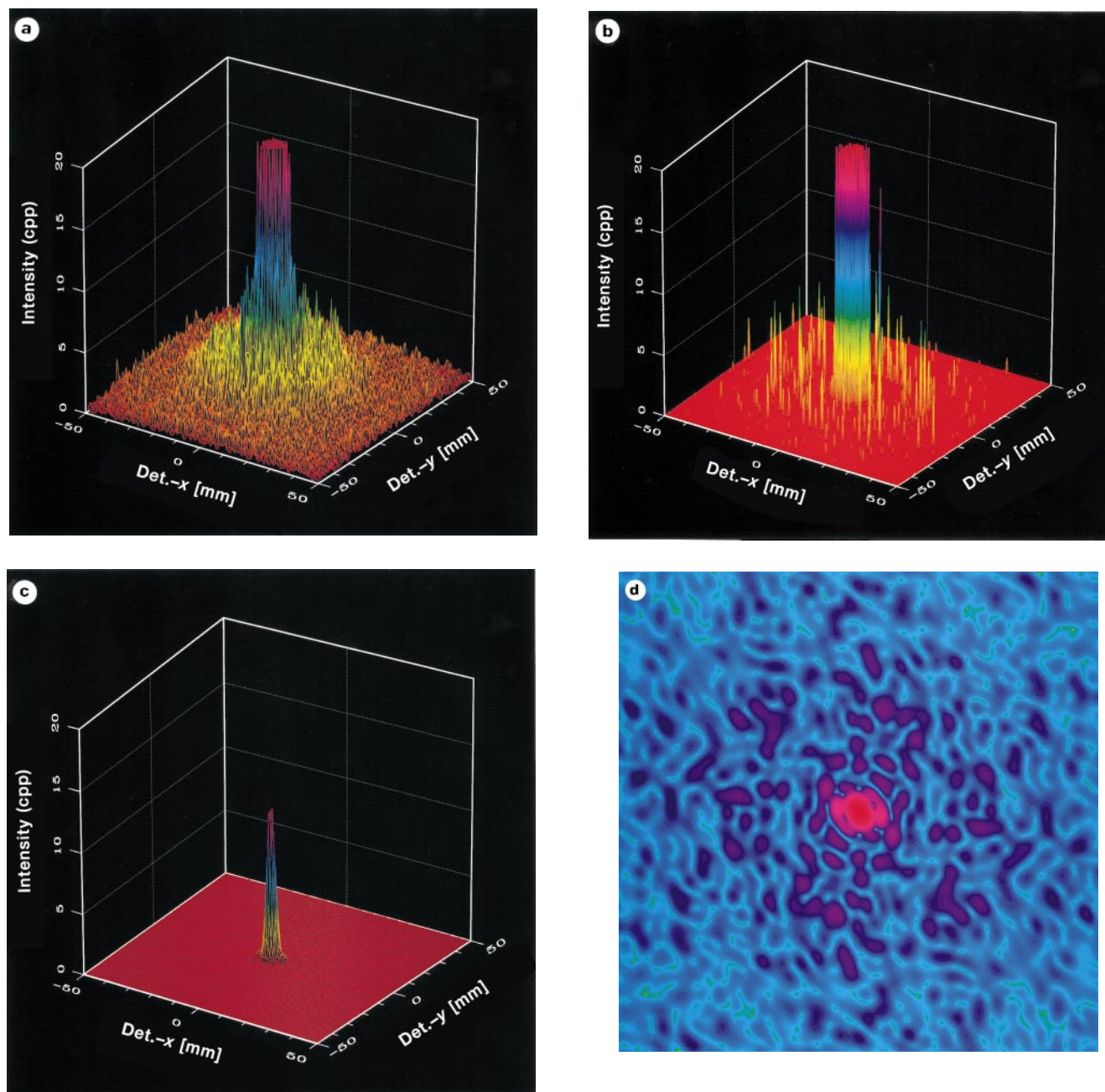


Figure 3 Elastic scattering from a variety of samples. **a–c**, Simulated diffraction images on a 128×128 pixel planar detector ($100 \text{ mm} \times 100 \text{ mm}$) normal to and centred at the beam, and placed 100 mm from the sample. The background was not modelled, and 100% detective quantum efficiency was assumed. The integrated X-ray intensity was 3×10^{12} (12 keV) photons per 100-nm diameter spot (3.8×10^6 photons per $\text{\AA}^2$); the pulse length was 10 fs . The resolution is $2.2 \text{ \AA}$ at the rim in **a–c**. cpp, counts per pixel. **a**, Scattering from a single tomato bushy stunt virus capsid. **b**, Scattering from a $5 \times 5 \times 5$ cluster of lysozyme molecules with an average r.m.s. conformational deviation

of $0.2 \text{ \AA}$ to model an imperfect lattice. **c**, Scattering from a single molecule of lysozyme. **d**, A planar section through the molecular transform (that is, a simulated continuous scattering image) of a single T4 lysozyme molecule under ideal conditions without sample movement or damage. Resolution at the rim of **d** corresponds to $2.0 \text{ \AA}$. Structure factor amplitudes are coloured logarithmically (magenta, high; green, low). The section is perpendicular to the z axis, and crosses through the origin at the centre of the image, revealing centric symmetry.

electrons will not be able to escape the increased positive potential of the sample. Trapped electrons will increase the kinetic energy of the sample through thermal equilibration, but they will also slow down the Coulomb explosion of the sample by partially neutralizing the positively charged protein core. These opposing effects have not been considered here, but are factors that are likely to influence the dynamics of larger systems.

In addition to the effects listed above, it has been suggested that extreme X-ray intensities could strip all outer-shell electrons from their parent atoms on a timescale of attoseconds⁵; however, a quantum mechanical analysis of the electric-field-induced tunneling shows that atoms actually become stabilized against ionization by this mechanism at high photon frequencies¹³ like those of X-rays.

We constructed a model in which X-ray-induced damage is described stochastically on the basis of the probability of a photoelectric or an inelastic event. The instantaneous probability of ionization of atom j at time t was calculated as the product of its photoelectric and inelastic cross-section^{10,11}, and the X-ray intensity $I(t)$. We modelled Auger emission as a stochastic exponential decay to reproduce appropriate K-hole lifetimes. The direction of photoemission was distributed according to a random deviate that followed a gaussian distribution⁸. A recoil velocity for the ionized atom owing to inelastic scattering or the emission of a photo- or Auger electron was determined from energy and momentum conservation. For inelastically scattered photons, the angle of deflection was determined by a random deviate following a Rayleigh distribution⁸. For each inelastic scattering the electron's recoil energy was calculated, and where this was greater than the binding energy of the electron an ionization event was modelled. An inventory was kept of what electrons remained on what atoms, and changes in the photoelectric, elastic and Compton scattering cross-sections of all atoms were computed and updated during exposures.

Energies and ionization events for a set of representative simulations on T4 lysozyme¹⁴ and its 118 crystallographically determined water molecules are shown in Fig. 1. The protein with its bound solvent molecules was considered to be in the gas phase, under conditions similar to a non-destructive electrospray mass spectrometry experiment (see below). For each simulation, a total flux of 3×10^{12} (12 keV) X-ray photons passed through the 100-nm diameter focal spot, corresponding to 3.8×10^6 photons per Å², and causing roughly 2,000 primary ionization events, or more than one ionization event for each non-hydrogen atom in the sample. For this set, the full-width at half-maximum (FWHM) for each pulse was

2 fs (Fig. 1a), 10 fs (Fig. 1b) and 50 fs (Fig. 1c). For an FWHM of 2 fs, only a quarter of the K-holes created by inner-shell ionization events had time to decay by Auger electron emission; therefore, the total number of positive charges on the sample at the completion of the 2-fs pulse was only 60% of the total number of charges at the completion of the 50-fs pulse. Furthermore, during short intense pulses numerous K-holes may be present at any one time, reducing the photoelectric cross-sections of atoms in which they were produced and thus lowering the total number of primary ionization events in the sample (see trend in Fig. 1). This effect makes the system radiation hardened to photo-ionization during very short exposures, and is more pronounced at higher radiation intensities (not shown).

The creation of a large number of positive charges in close proximity results in a rise in the electrostatic energy of the sample (Fig. 1), which drives its eventual explosion (Fig. 2). Each sample will survive only a single shot. The degree of conversion of potential energy into kinetic energy during the X-ray exposure is inertia limited, and so depends strongly on the duration of the pulse. During the 2-fs pulse there was insufficient time for the kinetic energy to grow appreciably (Fig. 1a). In contrast, by the completion of the 50-fs pulse the kinetic energy of the sample had surpassed its potential energy (Fig. 1c), indicating that the explosion of the sample was well under way. The destruction of the sample by the X-ray pulse is illustrated by snapshots from the trajectories for an FWHM of 2, 10 or 50 fs (Fig. 2a–c). For the two shorter X-ray pulse widths, only very small changes in the atomic positions have had time to develop.

Hydrogen ions and highly ionized sulphurs are the first to escape the immediate vicinity of the protein (at 12 keV the photoelectric cross-section for sulphur is about 50 times larger than that for carbon). In the 50-fs FWHM pulse, the molecule is destroyed before the pulse is over. The 2-fs FWHM simulation was continued to 50 fs beyond the pulse, and in this case the molecule explodes as well, but only after the structural information has been gathered. These results are in agreement with other observations. Molecular dynamics simulations of the response of small van der Waals clusters of atoms ionized by intense visible femtosecond laser pulses¹² also showed a delay of a few tens of femtoseconds before the creation of significant structural disorder, and when a femtosecond X-ray pulse generated by a laser plasma was used to probe the dynamics of a rapidly heated organic sample¹⁵ a delay of about 100 fs was observed before significant growth in disorder.

We calculated scattering intensities for a range of samples

Table 1 Calculated limits of resolution

Pulse duration*	1 fs	5 fs	10 fs	50 fs	100 fs	
Photons per pulse in a 100-nm diameter spot*	5×10^{13}	1×10^{13}	5×10^{12}	8×10^{11}	3×10^{11}	
Relative scattering power†	0.32	0.53	0.71	0.96	0.97	
Single lysozyme molecule	15 Å	24 Å	26 Å	30 Å	> 30 Å	1 photon per pixel‡§
	> 30 Å	> 30 Å	> 30 Å	> 30 Å	> 30 Å	9 photons per pixel‡
	682	223	150	33	12	Σphotons (2–30 Å)¶
2×2×2 cluster of lysozymes	2.5 Å	3.1 Å	4.8 Å	12 Å	17 Å	1 photon per pixel
	6.5 Å	12 Å	16 Å	30 Å	> 30 Å	9 photons per pixel
	5,795	1,901	1,277	277	105	Σphotons (2–30 Å)
3×3×3 cluster of lysozymes	< 2.0 Å	< 2.0 Å	< 2.0 Å	3.0 Å	6.5 Å	1 photon per pixel
	2.2 Å	3.0 Å	3.0 Å	12 Å	17 Å	9 photons per pixel
	17,254	5,661	3,803	824	313	Σphotons (2–30 Å)
5×5×5 cluster of lysozymes	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	1 photon per pixel
	< 2.0 Å	< 2.0 Å	< 2.0 Å	2.9 Å	3.9 Å	9 photons per pixel
	76,346	25,050	16,828	2,746	1,387	Σphotons (2–30 Å)
Single viral capsid (TBSV)	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	< 2.0 Å	1 photon per pixel
	< 2.0 Å	2.5 Å	2.5 Å	4.7 Å	22 Å	9 photons per pixel
	160,640	52,708	35,969	7,795	2,964	Σphotons (2–30 Å)

* Pairs of pulse lengths and integrated photon intensities were selected from the line of 'maximum tolerable damage' in Fig. 4b ($R_{\text{elec}} = 15\%$).

† Integrated scattering power of the sample over the exposure relative to the integrated scattering power of an undamaged, idealized sample.

‡ Two resolution limits are listed, corresponding to peak heights with 1 or 9 elastically scattered photons per pixel. Under ideal conditions, these values would correspond to intensity measurements with $I = 1\sigma(I)$ and $I = 3\sigma(I)$, respectively, where $\sigma(I)$ is the standard deviation of the intensity measurement. The background was not modelled, and 100% detective quantum efficiency was assumed. Wavelength = 1.0 Å (12 keV energy), planar detector (100 mm × 100 mm) normal to and centred at the beam, sample to detector distance = 100 mm, pixel size = 0.8 mm × 0.8 mm.

§ Resolution may be extended beyond this limit by averaging, if orientation can be determined accurately.

|| Information extending to this limit could be used for determining sample orientation.

¶ Number of elastically scattered photons on the detector between 2.0 and 30 Å resolution. An even larger number of photons will be scattered between 30 Å resolution and infinity (Fig. 3), but these are not listed.

(Fig. 3a–c), using pulse parameters from Fig. 1b. A single capsid of the tomato bushy stunt virus (TBSV)¹⁶ would scatter X-rays to atomic resolution with a 10-fs pulse containing 3×10^{12} (12 keV) photons in a 100-nm diameter spot (Fig. 3a). The capsid particle of TBSV is a $T = 3$ icosahedral nanocluster of 60×3 subunits. It is not a crystalline structure, as it lacks translational symmetry. For a tiny crystalline structure (Fig. 3b), high-resolution scattering may be obtained from a nanocrystal of $5 \times 5 \times 5$ lysozyme molecules arranged on a rectangular primitive lattice. Molecules in the nanocrystal and subunits in TBSV were given an average r.m.s. conformational deviation of 0.2 Å to simulate imperfect conditions. Although these structures diffract well with these pulse parameters, Fig. 3c shows that the same intensity would be scattered to only about 40 Å resolution (Fig. 3c) from a single lysozyme molecule. For comparison, Fig. 3d shows a planar section through the continuous molecular transform of lysozyme, shown here under ideal conditions without sample movement or damage. An analysis of maximum attainable resolutions in single exposures is given below (Table 1, Fig. 4).

Larger samples scatter to higher resolutions even without internal symmetry, and resolution for reproducible samples can be extended further by numerical alignment and averaging procedures^{17–19}. Such procedures are well established in electron cryomicroscopy. Averaging techniques exploit the fact that photon counts from the structural signal grow more rapidly than counts from an incoherent

background (sources of noise here include inelastic scattering, Bremsstrahlung from photoelectrons colliding with instrument walls, and an imperfect sample environment). Multi-image alignment and averaging methods developed in electron microscopy produced substantially increased resolutions, for example, for the ribosome¹⁹. With particles displaying high symmetry, resolution can be extended further by exploiting the symmetry of the structure^{17,20}; however, averaging procedures can only be applied when a reproducible sample scatters a sufficiently large number of photons for its orientation to be determined.

Phases for scattering images can be recovered in a number of ways, including the oversampling of continuous molecular transforms²¹, holographic imaging methods^{22,23}, holographic data evaluation methods²², classical methods of crystallography and techniques for phase extension from lower-resolution electron cryomicroscopy images.

The structural information, which is theoretically recoverable from the sample during an exposure, can be quantified by the introduction of a weighted-average agreement factor (R factor, see equation (2) in Methods), which provides a direct assessment of the data quality. The analysis is based on differences between scattering from a sample that suffers radiation damage and scattering from a hypothetical sample that suffers no radiation damage during X-ray exposure. Radiation damage interferes with the atomic scattering factors and positions. If the atomic scattering factors for the damaged sample are taken as unchanged from those of the parent atoms, then R provides information on the extent to which the average positions of the sample's nuclei are perturbed by X-ray-induced damage ($R \equiv R_{\text{nuc}}$). If the degree of ionization of each atom is included when modelling the atomic (or ionic) scattering factors, then R provides information on the extent to which the elastically scattered radiation is perturbed by X-ray-induced damage ($R \equiv R_{\text{elec}}$). R_{nuc} depends only on the movements of atoms during an exposure, whereas R_{elec} depends on both the movements and the changes in the electronic structure of atoms. Whereas R_{nuc} delivers useful insight into the explosion process, R_{elec} corresponds directly to the quality of the data that would be obtained in an actual experiment. Macromolecular crystal structures in the Protein Data Bank have crystallographic R factors of about 20%. Many of the structures, especially those collected earlier on photographic film, represent data sets with merging R factors in the 5% to 15% range. Taking the latter value as an arbitrary upper limit, we regard damage as acceptable if $R_{\text{elec}} \leq 15\%$. A survey of the landscape for damage tolerance (Fig. 4) shows many combinations of wavelength, integrated intensity and pulse length where $R_{\text{elec}} \leq 15\%$. Figure 4a and b shows contour plots of weighted-average R_{nuc} and R_{elec} values for 12 keV photons as functions of pulse duration and the total photon flux; R_{nuc} and R_{elec} display different response dynamics, as atomic positions do not change as fast as the electronic configuration of the atoms. Table 1 shows the calculated limits of resolution at five different points along the line of the 'maximum tolerable damage' in Fig. 4b. The total dose delivered to the sample at either end of this line represents an increase of several orders of magnitude above the previously postulated limit of about 200 photons per Å² for conventional experiments at low X-ray intensities and cryogenic temperatures². Increasing I_{tot} above 5×10^{13} photons per spot will probably be of little value even at very short pulse lengths, as above this dose most of the sample's electrons are stripped from the atoms during the X-ray exposure (see values for relative scattering powers in Table 1). Thus, the total number of elastically scattered photons no longer increases linearly with increasing dose, and X-rays scattered from unbound electrons will contribute additional noise. With increasing X-ray energy, the ratio of elastic-scattering events to damaging events becomes more favourable^{10,11}, and data quality improves as the X-ray probe moves to higher photon energies (Fig. 4c).

With ultrasmall samples, standard procedures for sample selec-

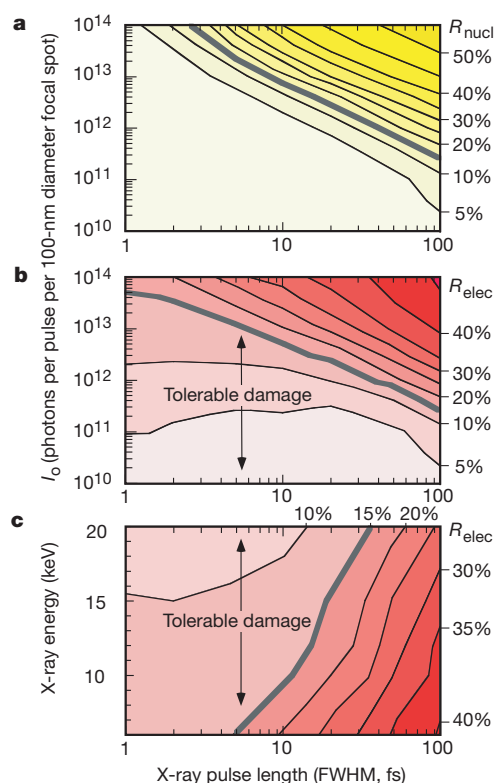


Figure 4 The landscape of damage tolerance. Contour plots of the nuclear (R_{nuc}) and electronic (R_{elec}) weighted-average R factors (equation (2)) as functions of the X-ray flux, pulse duration and photon energy made by interpolating the R value from the individual simulations. **a**, R_{nuc} for an X-ray beam of 12 keV, illustrating the extent to which the average positions of the sample's nuclei are perturbed by radiation damage. **b**, R_{elec} for a X-ray beam of 12 keV, illustrating the extent to which the information content of the elastically scattered X-rays is degraded by radiation damage. We regard damage as acceptable if $R_{\text{elec}} \leq 15\%$ which is indicated by the grey line. **c**, The variation of R_{elec} as the X-ray photon energy is changed, but with the total number of elastic-scattering events per carbon atom held constant. These intensities (I_{tot}) were 1.33×10^{12} (6 keV), 1.85×10^{12} (8 keV), 2.36×10^{12} (10 keV), 3.0×10^{12} (12 keV), 3.96×10^{12} (15 keV) and 6.0×10^{12} (20 keV) photons per 100-nm diameter spot.

tion and handling will no longer be applicable. New container-free methods are being developed²⁴ for characterizing, selecting and injecting single molecules, particles or nanoclusters into intense X-ray pulses. These methods are based on techniques used in electro-spray ionization mass spectrometers and cell sorters, and allow the injection of individual hydrated protein molecules, virus particles^{25,26} and intact ribosomes²⁶ into the vacuum chamber of the mass spectrometer while maintaining the structural integrity of the samples at cryogenic temperatures. The transit time of the sample is a few microseconds in the instrument. The sample entering the vacuum chamber is in random orientation, encapsulated in a micro-droplet. Such spraying methods could be used to characterize and inject well defined sample particles into intense X-ray pulses. Assembling protein molecules into nanoclusters (M. Svenda and J.H., manuscript in preparation) increases the intensity of scattered radiation from otherwise small proteins (Fig. 3). Such nanoclusters could be created by the specific attachment of target proteins on the surface of reproducible scaffolds, such as icosahedral viruses²⁷.

Should this new femtosecond window in imaging provide a path to high-resolution structural information without the need for macroscopic crystals, its impact on structural biology would be tremendous. □

Methods

Modelling

We created a program (XMD) that extends the GROMACS molecular dynamics package²⁸ to simulate electronic and structural changes triggered by X-rays in the sample. The GROMOS²⁹ force field was modified to incorporate Morse potentials for the description of all chemical bonds, thereby enabling bonds with sufficiently high energy to break. For water, the simple point charge model³⁰ was used and adapted in the same manner. Elastic, inelastic and photoelectric cross-sections of atoms were incorporated, and changes in the cross-sections during the X-ray exposure were modelled, using theoretical values^{10,11}. We did not model femtosecond collisional electron transfer and slower radical reactions. Modification of the dissociation energy of two atoms owing to electron emission was not taken into account.

We used T4 lysozyme¹⁴ as a model, including its 118 crystallographically determined water molecules. Hydrogens were added to polar and aromatic groups and to water oxygen atoms; the initial charge on the protein was +8. After several rounds of energy minimization the r.m.s. deviation from the crystal structure of the protein was 0.21 Å for all non-hydrogen protein atoms. This structure was used as the starting point for subsequent molecular dynamics simulations. Simulations were performed with a time step of 50 attoseconds, taking all non-bonded interactions explicitly into account.

Calculations

The X-ray pulse with intensity $I(t)$ was taken to have a gaussian temporal profile with pulse duration (FWHM) of 1, 2, 5, 10, 20, 50 or 100 fs. The total integrated X-ray flux, I_{tot} , was expressed as the number of 12-keV X-ray photons passing through a 100-nm diameter circular focal spot, and simulations were run with intensities I_{tot} of 10^{10} , 3×10^{10} , 10^{11} , 3×10^{11} , 10^{12} , 3×10^{12} , 10^{13} and 10^{14} photons for each of the pulse widths. A simulation with $I_{\text{tot}} = 0$ was the reference simulation for calculating elastic-scattering properties. The maximum intensity of the X-ray pulse was set at time $t = 0$, and the simulations were run from $t = -1.2$ FWHM to $t = +1.2$ FWHM (covering 99% of the integrated pulse). All simulations were repeated with different random number seeds for the stochastic events to test reproducibility. The results deviated by less than 5% from each other. We carried out further sets of simulations to estimate wavelength-dependent effects (see Fig. 4c, legend).

For unpolarized X-rays, the mean number of elastically scattered photons $I(\mathbf{u}, \Omega)$ to be detected by an idealized detector pixel of projected solid angle Ω centred at a positional vector $\mathbf{u}$ is

$$I(\mathbf{u}, \Omega) = 1/2(1 + \cos^2 2\theta) r_e^2 \int_{-\infty}^{\infty} I(t) \left| \sum_j f_j(t) \exp\{i\Delta \mathbf{k}(\mathbf{u}) \cdot \mathbf{x}_j(t)\} \right|^2 dt \quad (1)$$

where r_e is the classical electron radius (2.81785×10^{-5} Å); $I(t)$ is the intensity of the X-ray pulse; $f_j(t)$ is the atomic scattering factor for the j th atom as a function of time; $\mathbf{x}_j(t)$ is the position of this atom as a function of time; and $\Delta \mathbf{k}$ is the change in the wave vector of the X-ray photon when scattered through 2θ radians towards the pixel centred at $\mathbf{u}$. Radiation damage interferes with $f_j(t)$ and $\mathbf{x}_j(t)$. A direct assessment of imaging quality is given by the R-factor, defined as:

$$R = \sum_{\mathbf{u}} \left| \frac{K^{-1} \sqrt{I_{\text{real}}(\mathbf{u}, \Omega)} - \sqrt{I_{\text{ideal}}(\mathbf{u}, \Omega)}}{\sum_{\mathbf{u}} \sqrt{I_{\text{ideal}}(\mathbf{u}, \Omega)}} \right| \quad (2)$$

where

$$K = \frac{\sum_{\mathbf{u}} \sqrt{I_{\text{real}}(\mathbf{u}, \Omega)}}{\sum_{\mathbf{u}} \sqrt{I_{\text{ideal}}(\mathbf{u}, \Omega)}} \quad (3)$$

Scaling factor K describes the relative scattering power of the sample (Table 1). Values for I_{real} and I_{ideal} were computed from molecular dynamics simulations by evaluating equation (1) from 60 snapshots of each MD trajectory. I_{real} was derived from the time-dependent atomic coordinates $\mathbf{x}_j(t)$ and scattering factors $f_j(t)$ of a sample exploding in the X-ray pulse, while I_{ideal} was determined from the reference molecular dynamics simulation of an unexposed sample.

Received 23 December 1999; accepted 14 June 2000.

- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
- Henderson, R. Cryoprotection of protein crystals against radiation-damage in electron and X-ray diffraction. *Proc. R. Soc. Lond. B* **241**, 6–8 (1990).
- Solem, J. C. Imaging biological specimens with high-intensity soft X-rays. *J. Opt. Soc. Am. B* **3**, 1551–1565 (1986).
- Hajdu, J. in *Frontiers in Drug Research* (eds Jensen, B., Jorgensen, F. S. & Kofod, H.) 375–395 (Alfred Benzon Symposium 28, Munksgaard, Copenhagen, 1990).
- Doniach, S. Studies of structure of matter with photons from an X-ray free electron laser. *J. Synchr. Rad.* **3**, 260–267 (1996).
- Winick, H. The linac coherent light source (LCLS): A fourth-generation light source using the SLAC linac. *J. Elec. Spec. Rel. Phenom.* **75**, 1–8 (1995).
- Wiik, B. H. The TESLA project: an accelerator facility for basic science. *Nucl. Inst. Meth. Phys. Res. B* **398**, 1–8 (1997).
- Dyson, N. A. *X-Rays in Atomic and Nuclear Physics* (Longman, London, 1973).
- Krause, M. O. & Oliver, J. H. Natural width of atomic K and L level $K\alpha$ X-ray lines and several KLL Auger lines. *J. Phys. Chem. Ref. Data* **8**, 329–338 (1979).
- Veigele, W. J. Photon cross sections from 0.1 keV to 1 MeV for elements with $Z = 1$ to $Z = 94$. *Atomic Data* **5**, 51–111 (1973).
- Hubbell, J. H. et al. Atomic form factors, incoherent scattering functions, and photon scattering cross sections. *J. Phys. Chem. Ref. Data* **4**, 471–494 (1975).
- Ditmire, T. Simulation of exploding clusters ionized by high-intensity femtosecond laser pulses. *Phys. Rev. A* **57**, 4094–4097 (1998).
- Reiss, H. R. Physical basis for strong-field stabilization of atoms against ionization. *Laser Phys* **7**, 543–550 (1997).
- Weaver, L. H. & Matthews, B. W. Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. *J. Mol. Biol.* **193**, 189–199 (1987).
- Rischel, C. et al. Femtosecond time-resolved X-ray diffraction from laser-heated organic films. *Nature* **390**, 490–492 (1997).
- Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–372 (1978).
- Bottcher, B., Wynne, S. A. & Crowther, R. A. Determination of the fold of the core protein of hepatitis b virus by electron cryomicroscopy. *Nature* **386**, 88–91 (1997).
- van Heel, M. et al. Angular reconstruction in three-dimensional electron microscopy. Historical and theoretical aspects. *Scanning Microsc.* **11**, 195–210 (1997).
- Mueller, F. et al. The 3D arrangement of the 23 S and 5 S rRNA in the *Escherichia coli* 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 Å resolution. *J. Mol. Biol.* **298**, 35–59 (2000).
- Stowell, M. H. B., Miyazawa, A. & Unwin, N. Macromolecular structure determination by electron microscopy: new advances and recent results. *Curr. Opin. Struct. Biol.* **8**, 595–600 (1998).
- Miao, J. W., Charalambous, P., Kirz, J. & Sayre, D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature* **400**, 342–344 (1999).
- Szöke, A. Holographic methods in X-ray crystallography. 2. Detailed theory and connection to other methods of crystallography. *Acta Crystallogr. A* **49**, 853–866 (1993).
- Faigel, G. & Tegze, M. X-ray holography. *Rep. Progr. Phys.* **62**, 355–393 (1999).
- Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. Electrospray time of flight mass spectrometry of the intact MS2 virus capsid. *J. Am. Chem. Soc.* **122**, 3550–3551 (2000).
- Siuzyak, G. et al. Mass-spectrometry and viral analysis. *Chem. Biol.* **3**, 45–48 (1996).
- Rostom, A. A. Detection and characterization of intact ribosomes in the gas phase of a mass spectrometer. *Proc. Natl Acad. Sci. USA* **97**, 5185–5190 (2000).
- Lin, T. W., Porta, C., Lomonosoff, G. & Johnson, J. E. Structure-based design of peptide presentation on a viral surface: The crystal structure of a plant/animal virus chimera at 2.8 Å resolution. *Fold. Des.* **1**, 179–187 (1996).
- Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* **91**, 43–56 (1995).
- van Gunsteren, W. F. & Berendsen, H. J. C. *Gromos-87 Manual* (Biosmos BV, Groningen, 1987).
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. in *Intermolecular Forces* (ed. Pullman, B.) 331–342. (D. Reidel, Dordrecht, 1981).

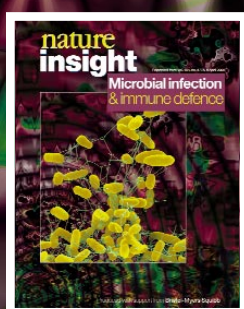
Acknowledgements

We are grateful to G. Faigel, S. Hövöller, A. Szöke, M. van Heel and A. Pratt for discussions. This work was supported by the Swedish Research Councils NFR and TFR as well as the EU-BIOTECH programme.

Correspondence and requests for materials should be addressed to J.H. (e-mail: janos@xray.bmc.uu.se).

nature insight

Microbial infection & immune defence



Cover illustration

Scanning electron micrograph of *Salmonella enteritidis*, rod-shaped bacteria that cause salmonellosis. (Image courtesy of Dennis Kunkel.)

Despite the extensive use of antibiotics and vaccination programmes, infectious diseases continue to be a leading cause of morbidity and mortality worldwide. Widespread antibiotic resistance, the emergence of new pathogens in addition to the resurgence of old ones, and the lack of effective new therapeutics exacerbate the problems.

The new approaches that are needed to deal with this increasing threat will come from the integration of two of the most active areas of biomedical research: the molecular and cellular basis of microbial pathogenesis, and the nature and manipulation of immune defence. In this month's Insight, we examine how bacteria attack and survive in the host, the mechanisms that the host uses to defend itself, and the therapeutic strategies that can be used to buttress these defences. As this is such a large and complex topic, we focus on bacterial disease; important health hazards such as HIV infection and malaria will be addressed in future Insights.

On page 760, Barry Bloom presents an overview of the topic, illustrating the complexity and diversity that arises from the particularity of microbial pathogens. The changing patterns of infectious disease globally and the evolving response strategies are discussed by Mitchell Cohen on page 762. Michael Donnenberg discusses on page 768 the myriad strategies by which enteric pathogens establish infection and survive in the face of an intense onslaught from the host immune response. On page 775, Chris Walsh extends this theme by examining the sophisticated methods that bacteria use to combat antibiotics. Walsh also describes new approaches to the development of novel antibiotics. Several exciting and recent findings have catapulted the innate immune system, which constitutes the first line of defence against infectious disease, from relative obscurity to the forefront of the fight against bacteria. Principal among these findings is the triggering of innate immunity Toll-like receptors on the surface of extracellular bacteria, a subject covered by Alan Aderem and Richard Ulevitch on page 782. Se-Ho Park and Albert Bendelac examine another recent surprise on page 788, namely the ability of mycobacterial lipids to trigger T-cell responses through presentation on non-classical class I molecules. Difficulties in generating vaccines against intracellular pathogens, and strategies to overcome these difficulties are explored on page 793 by Robert Seder and Adrian Hill. Finally, on page 799 Claire Fraser and co-workers update microbial genome sequencing and discuss how our increasing knowledge will provide new approaches to the diagnosis and treatment of infectious disease.

We are pleased to acknowledge the financial support of Bristol-Myers Squibb in producing this Insight. As always, though, *Nature* carries the sole responsibility for all editorial content and peer-review. We hope that both general readers as well as experts in the field will find these articles useful and informative.

Ursula Weiss Senior Editor

Ritu Dhand Insight Programme Editor

Publisher and liaison for corporate support **Liz Allen** (e.allen@nature.com)

On the particularity of pathogens

Barry R. Bloom

Harvard School of Public Health, Boston, Massachusetts 02115, USA

In the elemental struggle between pathogenic microbes and the immune system of the host, each strives for a unique advantage and thus each exploits its own unique particularities in pathogenesis and protection. And each presumably selects for the diversity that generally characterizes the wide range of successful host–pathogen interactions.

Almost two centuries ago, William Blake wrote: “He who would do good to another must do it in minute particulars”. On the basis of current knowledge, the contrary would seem also to be true for pathogens that cause disease, namely that they must do harm in minute particulars. The panoply of human and animal pathogens is large, enormously diverse, and constantly evolving, as we know from the continuing emergence of new infectious diseases and antibiotic resistance. Several years ago, Stanley Falkow outlined the requirements for any self-respecting pathogen as the ability to: enter the body; attain a unique niche; avoid, subvert or circumvent innate host defences; evade acquired specific immune responses; multiply or persist; cause tissue damage or disease (optional); and exit and transmit infection to new hosts. The scientific intrigue of infectious diseases relates to the need to learn the particularities of how each pathogen develops its unique strategy to survive within its host and assure transmission, how the host unleashes its multifarious immune mechanisms, and how the battle between pathogen and host leads to disease or recovery.

Evolution of diversity

Some examples illustrate the complexity and diversity of pathogens. Many think of *Escherichia coli* as the standard, harmless laboratory strain that is the host in which most molecular genetic manipulations are carried out. But at least eight different versions or pathotypes of *E. coli* produce quite different disease syndromes, depending on the nature of the virulence genes and toxins that they have picked up along their evolutionary pathway. There are versions that cause fluid loss and diarrhoea, haemorrhagic forms, kidney damaging forms and more, all evolved from a single species of pathogen. And whereas most Gram-negative-related pathogens such as *E. coli*, *Salmonella* and *Shigella* (the cause of bacterial dysentery) colonize the gut, they produce quite different diseases by attaching to different receptors or producing different toxins.

Cell-to-cell spread is a real challenge for pathogens that cannot succeed on their own and must multiply in host cells. *Listeria monocytogenes* and *Shigella dysenteriae* are both intracellular pathogens that have evolved a truly spectacular solution to the problem of spreading from cell to cell and causing enough tissue damage to be excreted and transmit infection. Each evolved a set of genes, curiously unrelated, to bind and polymerize host-cell actin molecules, thereby whirling the organisms about within the infected cell until by chance a few organisms are propelled like bullets into the cytoplasm of neighbouring cells.

From where did this extraordinary diversity and plasticity of pathogens derive? Some drug-resistant and antigenic

variants result from simple point mutations. More interestingly, many bacterial pathogens have exploited the most primitive forms of genetic information — DNA in plasmids and phages — to pick up whole chunks of genetic information, leap-frogging traditional slow evolution, to incorporate whole sets of genes from often unknown species in their environments. These acquisitions of borrowed genes for drug resistance or ‘pathogenicity islands’ provide fiendishly clever new opportunities for the pathogen to survive, outwit us, and produce disease. In one of the more spectacular instances, the principal determinants of virulence present in pathogenic but not in non-pathogenic strains of *Vibrio* (the bacterial agent of epidemic cholera) reside on a pathogenicity island that is actually a cryptic phage incorporated long ago into the genome. The phage encodes not only the gene encoding cholera toxin that is responsible for inducing the massive fluid loss in the gut which is characteristic of cholera, but also the receptor (toxin co-regulated pilus) for the phage which is delivered to host cells, just to be sure that there is no slip up in transmission.

Exploiting genome sequences

Perhaps the greatest advances in our understanding of infectious diseases and microbial pathogenesis over the next decade will derive from knowledge of the genomic DNA sequences of essentially all the major pathogens, which we may have by the end of 2000. Most pathogens have a lot of genes in common; yet each may have 30–40% of their genes found in no other organism, which may represent the particularity that makes each unique. Comparative genomics, for example between pathogens and related non-pathogens, will inform us about complex interactions between genes, virulence factors, and their environments.

We now live in a ‘post-genomic’ world in which every gene of every human pathogen can be displayed on the computer screen of every student and researcher. And we will shortly have the human and mouse genomes. The next step will be to learn what virulence determinants are expressed under conditions corresponding to the target niches in the host — the lung for tuberculosis, the gut for *Salmonella* — and to understand how the pathogens ‘sense’ their environments. Bacteria have extraordinarily subtle mechanisms for sensing the osmolarity of the environment, the nutrients surrounding them, the oxygen tension, and even quorum counts, that is, the density of bacteria in the neighbourhood. Knowledge of those patterns of gene expression will help define unique sets of regulated genes that can serve as important new targets for drugs.

The particularity of host responses to different pathogens is an equally challenging problem for understanding the specific necessary and sufficient immune

mechanisms required for resistance to each pathogen. We have learned a great deal about acquired immunity. The specificity of binding of antibodies to antigens has in many cases been elucidated at the atomic level. But many protective antigens are not linear peptide or carbohydrate epitopes, but rather three-dimensional structures, as seems to be the case for neutralizing epitopes of HIV, which cannot yet be mimicked by linear structures. We also know a great deal about the recognition by helper and cytotoxic T cells of precisely cleaved small peptides derived from protein antigens and presented in the context of class II and I antigens of the major histocompatibility complex. But we know far less about the mechanisms of natural immunity, those protective mechanisms that operate immediately upon infection. Here, as in the case of the evolution of microbial pathogenesis, we may have to borrow knowledge from our evolutionary predecessors. For example, Toll-like receptors are a common set of receptors that have been conserved for millions of years between the fruitfly *Drosophila melanogaster* and humans. They provide fruitflies with powerful resistance to microbial pathogens and are even able to discriminate between yeast and Gram-positive and -negative bacteria in both hosts, yet we have much to learn about whether the microbicidal mechanisms are similarly conserved.

Opportunities for new therapeutics

In this post-genomic era one might assume that the development of new generations of antibiotics and vaccines to stem the emergence of drug resistance for virtually every microbial pathogen is inevitable. Indeed the opportunities are almost limitless for identifying new microbial targets for rationally designed drugs that are unique and selective for the pathogen, minimizing the risks of untoward adverse effects on the host. With robotic high-throughput drug-screening technology and the power of combinatorial chemistry to produce

more compounds in a week than the largest drug company in the world could have in a year, the scientific possibilities for antimicrobial drug development are similarly promising. But the principal limitations on development of new drugs and vaccines may not ultimately be scientific, but practical. It is the particularity of each pathogen that imposes enormous costs and weighty decisions on industry. Will broad-spectrum or cross-cutting drugs that could stem the epidemic of drug-resistant tuberculosis, for example, be developed only if they can assure returns on investment because they also act on common bacterial infections like Streptococci or Staphylococci? Or are there possible niche markets for drugs that take advantage of the particularities of individual pathogens and are highly selective in their activity, and are therefore less likely to contribute to the general rising level of antibiotic resistance of unrelated pathogens? Can one produce simple vaccines against pathogens with multiple antigenic types or must the vaccine encompass every possible antigenic type? With a decade required to develop a new drug, at a cost of US\$300–500 million, these will be difficult resource allocation decisions, even in the context of the industrialized world. They become excruciating when one recognizes that the major burden of infectious diseases falls on people in developing countries who have the fewest resources to create markets for new products.

An epistemological principle of science is that from an understanding of the particularities of complex systems will emerge some general and simple truths. It was Oliver Wendell Holmes who said: “For the simplicity on this side of complexity, I care not a fig. For the simplicity on the other side of complexity, I would give my life.” It is unlikely that many microbiologists will soon be going over the cliff, because they are consumed at the moment by wonder and fascination at the richness of the particularities — and potentialities — of microbe biology.

Changing patterns of infectious disease

Mitchell L. Cohen

Division of Bacterial and Mycotic Diseases (C09), National Center for Infectious Diseases, Centers for Disease Control and Prevention (CDC), 1600 Clifton Road, Atlanta, Georgia 30333, USA

Despite a century of often successful prevention and control efforts, infectious diseases remain an important global problem in public health, causing over 13 million deaths each year. Changes in society, technology and the microorganisms themselves are contributing to the emergence of new diseases, the re-emergence of diseases once controlled, and to the development of antimicrobial resistance. Two areas of special concern in the twenty-first century are food-borne disease and antimicrobial resistance. The effective control of infectious diseases in the new millennium will require effective public health infrastructures that will rapidly recognize and respond to them and will prevent emerging problems.

For most of the twentieth century, the predominant feeling about the treatment, control and prevention of infectious diseases was optimism. In 1931, Henry Sigerist wrote¹, 'Most of the infectious diseases ... have now yielded up their secrets.... Many illnesses ... had been completely exterminated; others had [been brought] largely under control....' Between 1940 and 1960, the development and successes of antibiotics and immunizations added to this optimism, and in 1969, Surgeon General William H. Stewart² told the United States Congress that it was time to 'close the book on infectious diseases'. With victory declared, increasing emphasis was directed at the non-infectious diseases such as cancer and heart disease. Often, research on infectious disease or activities on their prevention and control were de-emphasized and resources were reduced or eliminated. As recently as the 1980s, pharmaceutical companies, believing that there were already enough antibiotics, began reducing the development of new drugs or redirecting it away from antibiotics^{3,4}. In much of the developed world, the public had forgotten the impact of infectious diseases on previous generations and shared in the confidence that modern medicine and technology would prevail.

This optimism was soon shaken by a series of outbreaks and epidemics of new, re-emerging and antimicrobial-resistant infections. Legionnaire's disease, Ebola virus, HIV, 'flesh-eating' bacteria and 'mad cow disease' (bovine spongiform encephalopathy) were among the topics that began to appear in both scientific journals and the popular press. These diseases, occurring both in the developing and developed worlds, indicated that much was still unknown about infectious diseases. At the beginning of the twenty-first century, infectious diseases were once again capturing the attention of public health workers, academics, government and the general public. Here I will review how infectious diseases changed during the twentieth century, why infectious diseases have been emerging, and what will be necessary to address this important public health problem in the twenty-first century.

Infectious diseases in the twentieth century

At the beginning of the twentieth century, infectious diseases were the leading cause of death worldwide. In the

United States, three diseases — tuberculosis, pneumonia, and diarrhoeal disease — caused 30% of deaths⁵ (Fig. 1a). The average life expectancy was 47, in large part because of the high infant and childhood mortality from infection. A child born in 1900 had almost a 10% chance of dying between the ages of one and four, often from pneumonia or diarrhoeal disease⁶. In the autobiography of his depression-era childhood, *Angela's Ashes*, Frank McCourt described the devastation of losing three of six siblings to infections and almost dying of typhoid fever himself. This experience was not unusual for much of the world.

Surprisingly, by 1900, morbidity and mortality from infectious diseases had already considerably improved in much of the developed world. Between 1700 and 1900, the average life expectancy in Britain had increased from 17 to 52 years⁷, and the death rate from tuberculosis had fallen by 80% (ref. 8). The improvement in life span and decreases in mortality from infectious disease were attributed to a series of factors that were decreasing host susceptibility and curtailing disease transmission: better nutrition and housing, safer food and water, and improved hygiene and sanitation (Fig. 2).

Although deaths from many infectious diseases were already declining, the introduction of antimicrobial agents in the mid-twentieth century accelerated these declines even further. In England and Wales, deaths from childhood fever, caused by *Streptococcus pyogenes*, fell by more than 50% after the introduction of sulphadiazine⁹. In the United States, deaths from infectious disease declined at an annual rate of 8.2% between 1938 and 1952; the most prominent decreases were in tuberculosis and pneumonia¹⁰. These decreases coincided with the beginning of the era of antibiotics. The influence of immunization was also striking: in 1952, 57,879 cases of paralytic poliomyelitis were reported in the United States; by 1965, there were 72 (ref. 11). The eradication of smallpox in 1977, also a triumph of immunization, was one of the greatest accomplishments of public health. Between 1900 and 1980, mortality from infectious disease fell from 797 to 36 per 100,000 (ref. 10). By the end of the twentieth century, in most of the developed world, mortality from infectious diseases had been replaced by mortality from chronic illnesses such as heart disease, cancer and stroke (Fig. 1b). In 1997 these three conditions caused 62% of all deaths in

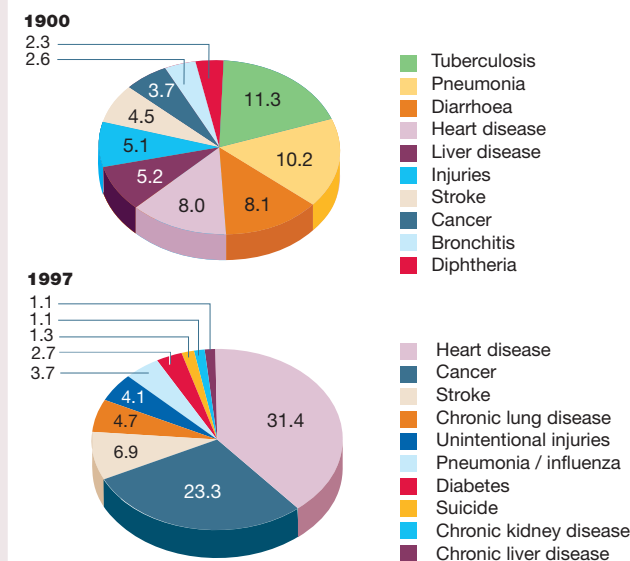


Figure 1 The ten leading causes of death in the United States in 1900 and 1997. Infectious diseases that were the most important causes of death at the beginning of the twentieth century have been replaced by chronic diseases.

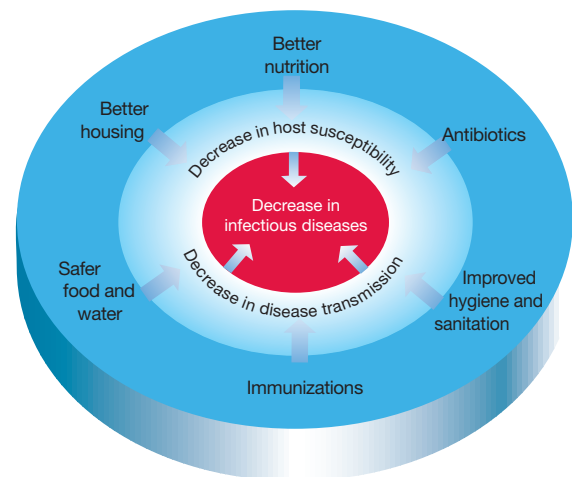


Figure 2 Factors influencing the decrease in infectious diseases in the twentieth century. Most factors led to decreases in host susceptibility and/or disease transmission.

the United States. Between 1900 and 1997, the average life span had increased by about 60% to more than 76 years¹².

The developing world had not had the same success with infectious diseases: there they remained the major cause of morbidity and mortality. In 1998, the World Health Organization estimated that infectious diseases caused over 13 million deaths — almost a quarter of the 54 million deaths worldwide¹³. Among the most common causes of mortality were the three diseases that had been so common in the developed world at the beginning of the twentieth century: pneumonia (3.5 million), diarrhoeal disease (2.2 million) and tuberculosis (1.5 million). Other important infectious causes of death were AIDS (2.3 million), malaria (~1.1 million) and measles (1.0 million). Many of these deaths, particularly from pneumonia and diarrhoeal disease, occurred in small children. The societal and technological advances that had influenced infectious diseases in the developed world had less of an effect in the developing world.

At the end of the twentieth century there were worrying trends in both the developed and the developing worlds. New infectious diseases and microorganisms were being recognized: Legionnaire's disease, toxic shock syndrome, Lyme disease, HIV, Nipah virus, hantavirus, *Escherichia coli* O157:H7, 'flesh-eating' bacteria, and many others. Infectious diseases were also being recognized as the cause of chronic illnesses; *Helicobacter pylori*, for instance, is now known to be the cause of peptic ulcers. New infectious agents, such as Ebola or Marburg virus, had the potential for rapid international spread. Diseases such as cholera, tuberculosis, dengue fever, yellow fever and malaria, which had once been controlled in many parts of the world, were re-emerging. Resistance to antimicrobial agents was becoming a serious global problem. Even in the developed world, infectious disease mortality was increasing. In the United States from 1981 to 1995, this increase was at a rate of 4.8% per year from 36 to 63 deaths per 100,000 (ref. 10).

In 1992, the Institute of Medicine (IOM), recognizing these trends, published a report¹⁴ entitled *Emerging Infections: Microbial Threats to Health in the United States*. This report defined an emerging infection as a "new, reemerging, or drug-resistant infection whose incidence has increased in the last two decades, or whose incidence threatens to increase". The report examined this emergence,

identified factors that were influencing it, and suggested approaches to addressing the problem.

Factors influencing emergence

The IOM report identified six factors (Fig. 3) as influencing the emergence of infectious diseases. Many of these factors increase the susceptibility of populations to infectious diseases or increase the exposure to or transmission of infectious agents. Emergence is often the consequence of societal and technological change and is frequently unexpected and unpredictable. In most instances the emergence of a specific agent results from a complex interaction of several factors that can vary even by geographic area. For example, vancomycin-resistant enterococci (VRE) emerged in hospitals in the United States because of antibiotic use combined with inadequate infection control practices and simultaneous increases in the number of susceptible persons in intensive-care units. In Europe, the emergence of VRE might have also been influenced by the agricultural uses of avoparcin, a glycopeptide antibiotic used in Europe as an animal growth promoter. Ironically, some factors that have resulted in a decline in one disease can contribute to an increase in another. The development of refrigeration, for instance, made food safer by inhibiting the growth of most food-borne pathogens, but it provided an advantage to organisms such as *Listeria* or *Yersinia* that can grow in the cold.

The recurring theme throughout all of these factors that influence the emergence of infectious diseases is change. During the twentieth century, tremendous societal and technological changes occurred, and these are likely to continue or accelerate in the twenty-first century. It is useful to examine how these changes have influenced infectious disease emergence so that we can anticipate some of the challenges of the new century.

Changes in demographics and behaviour

Demographic changes fall into several broad areas: changes in population, such as the increasing prevalence of persons with susceptibility to infection; societal changes, such as increases in two-income or single-parent families (which in turn lead to a greater use of day care for children and subsequently to an increase in disease transmission); and movements of infected or high-risk populations

by immigration. Changes in behaviour include various factors from increases in recognized risky behaviours, such as unsafe sex or the use of alcohol or drugs, to risks that are often unrecognized, such as recreational activities, exposures to new geographic areas and patterns of use of antimicrobial agents.

One of most important demographic factors influencing the emergence of infectious diseases has been the increase in susceptible populations. Ageing of the population, increases in underlying diseases, and technological advances in health care have all contributed. At the beginning of the twentieth century, for instance, less than 5% of the US population was over 65; by 2040 it is estimated that more than 25% will be that age or older¹⁵. Ageing contributes to susceptibility to many bacterial diseases, even in the absence of other underlying conditions. Thus, with an ageing population, there are more persons that are susceptible to certain diseases and a larger group that might sustain transmission. Similar increases are occurring in the number of persons with underlying diseases or conditions that contribute to susceptibility. Malnutrition contributes to susceptibility to infectious diseases in many parts of the developing world. The HIV epidemic in Africa has produced an enormous population of susceptible persons¹⁶. Other diseases such as diabetes, autoimmune diseases, malignancies and renal failure can also suppress host defences and increase the likelihood of infection. The prevalence of many of these chronic conditions is increasing, in part because the life span of patients can now be prolonged. Between 1935 and 1995 the prevalence of diabetes in the US population increased from 0.5% to 3% (ref. 17). This is directly related to causative factors such as obesity but also to improvements in medical technology, including the availability of insulin. It is estimated that, including undiagnosed cases, there could be more than 16 million persons with diabetes in the United States.

Medical technology is also increasing the survival of persons with other conditions that might render them susceptible to infection. Between 1960 and 1980, effective chemotherapy increased the five-year survival for persons with Hodgkin's disease from 40% to 76% (ref. 18). Persons with malignancies are often at risk of infections during chemotherapy, but some might have lifelong susceptibility to certain infections even after the successful treatment of their malignancy. Improvements in the technology of intensive care units (ICUs) and of organ transplantation have increased their use but have also caused the numbers of susceptible persons to increase and have provided opportunities for the emergence of new diseases. The increase in ICU patients has contributed to both the emergence of antimicrobial resistance and a greater incidence of fungal diseases. Debilitated patients, multiple courses of powerful antimicrobial drugs, long periods in hospital and often inadequate practices for controlling infection are factors in the emergence of methicillin-resistant staphylococci, VRE and a variety of multidrug-resistant Gram-negative rods. Some of the same factors have contributed to the emergence of fungal diseases, such as disseminated *Candida* infections, which in the 1980s increased 11-fold in hospital patients in the United States¹⁹. Organ transplantation has created one of the most susceptible populations and consequently has brought about increases in infections with *Listeria*, mycobacteria, viral agents and moulds²⁰. Some of the moulds, such as *Aspergillus* spp. and fungi with low pathogenicity, were uncommon or previously unrecognized as causes of human disease.

Human behaviour has provided more opportunities for exposure to and transmission of new and re-emerging infection and have increased selective pressure for antimicrobial resistance. Sexual activity is an important behavioural risk factor that has contributed to emergence. The rapid spread of HIV in heterosexual populations in parts of Africa was facilitated by social and economic factors that encouraged multiple sexual partners¹⁶. In the western world, the sexual revolution that began in the 1960s was associated with increases in sexually transmitted diseases. The appropriate and inappropriate use of antimicrobial drugs to treat or prevent the increasing number

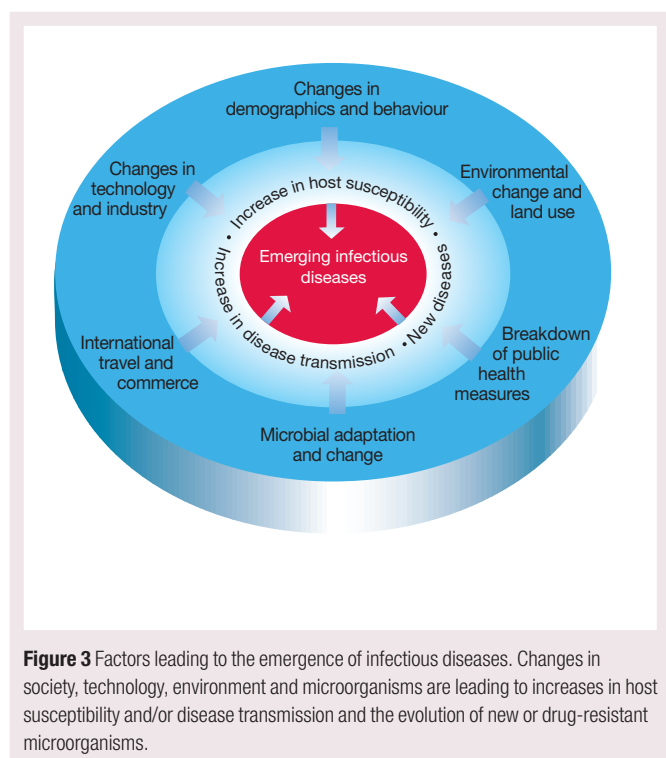


Figure 3 Factors leading to the emergence of infectious diseases. Changes in society, technology, environment and microorganisms are leading to increases in host susceptibility and/or disease transmission and the evolution of new or drug-resistant microorganisms.

of sexually transmitted infections contributed to the emergence of drug resistance in gonorrhoea and chancroid. Other behaviours are also risk factors for emergence. Smoking has been associated with pneumococcal disease²¹. Intravenous drug use has been associated with sexually transmitted diseases, including HIV²². Changing eating habits are exposing people to new foods with unfamiliar risks, to foods eaten without further preparation or to foods prepared outside the home. Thus, the increase in spending on food prepared or consumed outside the home between 1970 and 1990 parallels the increase in the percentage of food-borne outbreaks occurring outside the home²³. Recreational activities or moves to new geographic areas can also pose new risks. Spelunking (potholing) has been associated with outbreaks of histoplasmosis; another mycosis, coccidioidomycosis, has increased in part because previously unexposed persons moved or retired to areas in the southwestern United States where the fungus is endemic²⁴.

One of the strongest influences on the emergence of resistance has been the unnecessary use of antimicrobial agents. In many parts of the world, antibiotics are available over the counter. Even in countries in which use is more controlled, their use is often unnecessary²⁵. Many of the antimicrobial drugs prescribed to outpatients are for the treatment of upper-respiratory infections. Because most of these infections are viral, the antimicrobial agents are ineffective and unnecessary. In 1992, 18% of the 110 million courses of outpatient antibiotics prescribed in the United States were for upper-respiratory infections, providing strong selective pressure for the emergence of resistance in bacterial pathogens such as *Streptococcus pneumoniae*²⁶.

Changes in technology and industry

The twentieth century witnessed a number of technological advances that improved life and health and often transformed industry. However, some of these technologies were unexpectedly associated with the emergence of disease. Air-conditioning cooling towers were linked to Legionnaire's disease, super-absorbent tampons to toxic shock syndrome, and the fast-food hamburger to *E. coli* O157:H7. In the food industry especially, tremendous changes have occurred in how food is produced, preserved and processed. Antimicrobial agents used for promoting growth and preventing and treating disease in animals have facilitated large-scale, lower-cost production

but have contributed to the emergence of drug resistance in *Salmonella* and *Campylobacter* that are transmitted through the food chain to humans. Changes in rendering and in the feeding of rendered materials to food animals contributed to the emergence of 'mad cow disease'. There has been an increased emphasis on foods with a longer shelf life or more natural foods preserved only by refrigeration. This latter phenomenon has provided conditions favourable to *Listeria*, an organism that grows at refrigeration temperatures. The food industry has changed drastically, with consolidations and mergers leading to a broad geographic distribution of products. Thus, the typical food-borne outbreak vehicle that was once the home-made potato salad at the local church supper has now become any one of a number of commercial products distributed in multiple states or countries²⁷.

Environmental change and land use

Changes in environment and in land use are global activities that include both natural and man-made changes. In the developing world, these include the encroachment on the tropical rain forest, which poses a risk for the emergence of new haemorrhagic fever viruses. They also include the impact of growing megapolises often with inadequate hygiene and sanitation. As pointed out by Laurie Garrett, author of *The Coming Plague*²⁸, these cities are time bombs for the emergence and transmission of infectious diseases. The spread of cholera in Latin America during the 1990s was facilitated by such conditions. In the developed world, agricultural practices have contributed to blooms of *Pfiesteria* and toxic algae. Conservation efforts, such as preserving deer populations, have led to the spread of Lyme disease. Climate change, whether natural or man-made, is also contributing to the emergence of infectious diseases. In 1993, in the southwestern United States, increased rainfall led to increased vegetation, larger rodent populations, more rodent-human contact and the first recognized outbreak of hantavirus in North America²⁹. Similarly, in 1997 higher water temperatures in the Pacific Northwest, caused by El Niño, provided unusual conditions that were favourable for the growth of *Vibrio parahaemolyticus* and resulted in shellfish-associated outbreaks of disease³⁰.

International travel and commerce

One of the most impressive changes of the twentieth century has been in the ease of international travel. Travel that once required months has been reduced to hours, converting the world into a global village. Such travel facilitates the global transmission of diseases spread by person-to-person contact such as shigellosis or gonorrhoea. Respiratory infections, too, can spread more rapidly, as demonstrated by the outbreaks of influenza associated with Alaskan cruises³¹. The recent outbreak of a vector-borne agent, West Nile virus, in the northeastern United States could have been caused by the introduction of an infected bird or person³². Specific strains of drug-resistant *Strep. pneumoniae* or *Neisseria gonorrhoeae* have been traced between continents, suggesting that they were transmitted by human travel. Diseases that would normally have a limited geographic distribution have become part of the differential diagnosis of the unwell traveller. The febrile emergency-room patient in the local hospital can easily be a refugee or an immigrant or even a tourist returning from Africa with malaria or schistosomiasis. As with infected travellers, contaminated food can also move rapidly from continent to continent. International food-borne disease outbreaks have become more common and have introduced new pathogens to an area, such as the recent outbreaks in the United States of *Cyclospora* associated with imported raspberries³³.

Microbial adaptation and change

As society and technology change, so do microorganisms. In some instances, change is the result of selective pressures, such as the use of antimicrobial agents; at other times the cause of microbial change is less clear. A recent example of microbial evolution is the emergence of

E. coli O157:H7. This bacterium is an important cause of food-borne disease, causing bloody diarrhoea and a potentially fatal illness called the haemolytic uraemic syndrome. It is likely that this organism emerged since 1950, after the acquisition of *Shigella* toxin genes by *E. coli* O55 (ref. 34). This genetic event resulted in a new pathogen that combined the low infectious dose and toxigenicity of *Shigella* with the resistance to adverse environmental conditions that is characteristic of *Salmonella*. This combination of characteristics, coupled with changes in the food industry, are probable explanations for this organism's success.

Breakdown of public health measures

Some infectious diseases have emerged or re-emerged because the public health systems established to prevent or control them have broken down. A variety of elements have contributed to this breakdown, including complacency from past successes against infectious diseases, limited resources and competing priorities in public health, social unrest, wars, and population movements. The global re-emergence of tuberculosis is a good example of several of these factors³⁵. In the United States in the early 1980s, tuberculosis was targeted for elimination by the early twenty-first century. The decline in incidence had been consistent and resembled a straight line reaching zero in 2010. In some public health jurisdictions, tuberculosis was rare, and control programmes faced the competing priorities of HIV and chronic diseases. Tuberculosis treatment programmes to ensure compliance and decrease the emergence of drug resistance such as directly observed therapy (DOT) were curtailed. What was not realized was the impact of HIV or immigration on tuberculosis, which soon led to a re-emergence of the disease. In addition, because programmes such as DOT were no longer ensuring appropriate treatment, the cases of tuberculosis that re-emerged were often resistant to multiple antituberculosis drugs. A simultaneous emergence of tuberculosis has occurred in parts of the developing world that have been most affected by the HIV epidemic. In this instance, the often intrinsically high rates of tuberculosis infection, coupled with the immunosuppression of HIV, have resulted in substantial increases in active disease. In the countries of sub-Saharan Africa, annual rates of tuberculosis have increased by as much as 15% (ref. 36).

Emerging infections in the twenty-first century

As was pointed out by the historian George Santayana in *The Life of Reason*, "those who cannot remember the past are condemned to repeat it". The twentieth century was a microcosm of the history of infectious diseases. During one century, most of the developed world experienced vast improvements in health and in the prevention and control of infectious diseases at rates that dwarfed previous centuries. However, much of the improvement was limited in the developing world; both the developed and developing world experienced important health problems in the emergence of new and once-controlled infections. Societal and technological change accounted for both the control and the emergence of infectious diseases. It is likely that such change will continue into the twenty-first century and that the rate of change might accelerate. One lesson learned from history is that change leads to the continued emergence of infectious diseases, and we must be prepared to address this problem. We can never be too complacent about infectious diseases, or we will find ourselves in circumstances as threatening as the re-emergence of multidrug-resistant tuberculosis in the late 1980s or the influenza epidemic of 1918 in which 20–25 million persons perished worldwide¹⁰.

It is likely that researchers in the twenty-first century will identify infectious agents as contributors to many chronic diseases. The infection itself might be symptomatic or asymptomatic. Viral agents, for instance, now are being recognized as causes or important cofactors in a number of cancers. *Chlamydia* or other infections might have a role in coronary heart disease. Ulcers have been linked to *H. pylori* infections. The haemolytic uraemic syndrome has been associated with *E. coli* O157:H7 infections. Guillain-Barré syndrome can follow

Campylobacter infections, and a variety of enteric infections can lead to reactive arthritis or Reiter's syndrome. With a better understanding of both pathogen and host response, the new technologies of the twenty-first century will lead to diagnostic tools that will clarify the role of microorganisms in many chronic diseases.

Two areas deserve particular attention as we begin the new century: food-borne diseases and resistance to antimicrobial agents.

Food-borne disease

Food-borne disease was recognized as a high priority by the IOM¹⁴: "The potential for foods to be involved in the emergence or re-emergence of microbial threats to health is high, in large part because there are many points at which food safety can be compromised."

At the beginning of the twentieth century there were many challenges to food safety, including diseased animals, unsafe food handling, inadequate food preservation, unsafe water and poor sanitation, and the consumption of raw foods such as milk and shellfish. Food-borne outbreaks of typhoid, scarlet fever, cholera, hepatitis, staphylococcal food poisoning and brucellosis were common. Poor nutrition also contributed to the general susceptibility to infection. The incidence of these food-borne diseases was soon reduced by a series of regulatory and technological changes, including food inspection, hygienic processing, refrigeration, safe canning, food additives and preservatives, and pasteurization. Nutrition was improved by the fortification of food with vitamins and minerals, and social programmes to feed the poor. However, although many of the classic food-borne diseases were on the wane, new diseases were emerging²³. The technological and societal factors discussed above led to the emergence of non-typhoid salmonellosis, campylobacteriosis, listeriosis, and infections involving *E. coli* O157:H7, *Cyclospora*, calicivirus and *Vibrio vulnificus*. It has been estimated that food-borne illness accounts for 76 million illnesses, 325,000 periods in hospital and 5,000 deaths in the United States annually³⁷.

In earlier generations, the malnutrition from inadequate amounts or poor quality of food increased people's susceptibility to infections. It is ironic that for present generations, malnutrition from excessive amounts or types of food is increasing susceptibility to infectious diseases by contributing to heart disease, diabetes, hypertension and obesity. In addition to the increases in susceptible populations, as pointed out in the IOM report, there are many points at which the safety of food can be compromised. In the twenty-first century, international commerce and markets will continue to increase, new technologies will affect food production, processing and preservation, and the public will eat more ready-to-eat foods and more meals outside the home. We are increasingly dependent on others to ensure the safety of our food.

Resistance to antimicrobial agents

Antimicrobial resistance was also targeted by the IOM¹⁴: "Microbes that once were easily controlled by antimicrobial drugs are, more and more often, causing infections that no longer respond to treatment with these drugs". Antibiotics were one of the great discoveries of the twentieth century. However, the emergence of antimicrobial resistance was recognized soon after the discovery of penicillin and has followed the introduction of most every new drug. The problem has greatly escalated as we enter the twenty-first century³⁸ and a series of microbes that present serious challenges are listed in Table 1. A combination of increased selective pressure from the use of antimicrobials, increased disease transmission and a decline in the development of new antibiotics has raised the spectre of once-treatable infections becoming untreatable. In the hospital, one of the first serious problems was the emergence of VRE. Between 1989 and early 1999, the percentage of VRE isolated from patients in hospital in intensive-care units in the United States increased from 0.4% to 25.9% (ref. 39). Although the enterococcus is less virulent than many other hospital-acquired pathogens, VRE infections were essentially untreatable until the introduction of quinupristin/dalfopristin. A

potentially more serious problem is emerging with *Staphylococcus aureus*. Strains already resistant to methicillin are becoming resistant to vancomycin, often the drug of last resort. Because in many hospitals more than 40% of *Staph. aureus* strains are resistant to methicillin, there are ample opportunities for the emergence of vancomycin-resistant strains³⁹. At the beginning of the twenty-first century, at least five strains with intermediate resistance to vancomycin have been reported in the United States and Japan⁴⁰. Other drug-resistant bacteria are also posing problems for hospitals. Various Gram-negative bacteria are becoming increasingly resistant to extended-spectrum cephalosporins, some of the newer and most powerful antimicrobial agents.

Resistance to antimicrobial drugs is not just a problem in the hospital environment. Communities have also experienced problems in both the developing and developed worlds. In the 1970s, drug-resistant *N. gonorrhoeae* and *Haemophilus influenzae* were recognized worldwide. Although other antimicrobial agents were available for those infections, parts of the developing world have experienced outbreaks of drug-resistant bacteria such as *Shigella dysenteriae* for which alternative antimicrobial therapy was often not available⁴¹. Drug resistance in the developed world has been increasing in food-borne pathogens such as salmonellae and *Campylobacter* and has been attributed to the use of antimicrobials in food animals. Between 1979 and 1994 the frequency of multiple drug resistance in *Salmonella* increased from 17% to 31% (ref. 42). Between 1991 and 1999, ciprofloxacin resistance in *Campylobacter* increased from 0% to 13.6% (ref. 43).

One of the most worrying problems for the community has been the emergence of drug-resistant *Strep. pneumoniae* (DRSP). This poses a serious problem for both the developed world and the developing world. In the developing world this bacterium is a major cause of pneumonia and death in both children and adults. In the United States it causes two million cases of pneumonia and seven million middle-ear infections. The frequency of penicillin resistance varies between hospitals and geographic areas, but between 1993 and 1997 it increased in the United States from 14% to 25% (ref. 44). The increase in antimicrobial resistance has led to the emergence of strains that are susceptible only to vancomycin and has already resulted in changes in the recommended empirical treatment of meningitis in children to include vancomycin. The emergence of antimicrobial resistance has not been limited to bacteria: it is an important public health issue for fungi such as *Candida*, viruses such as HIV, and parasites such as malaria. Even if new antimicrobial agents are introduced, their use will provide selective pressure for the emergence or resistance of these strains. It is likely that antimicrobial resistance will be a serious problem for the twenty-first century.

Addressing emerging infections in the twenty-first century

As the conditions encouraging the emergence of infectious diseases are likely to persist into the twenty-first century, it is necessary to develop flexible strategies to detect and respond to such problems rapidly. Several national and international health organizations have developed plans to address emerging infectious diseases. One plan developed by the Centers for Disease Control and Prevention identifies four major strategies: (1) enhancing surveillance and response, (2) encouraging applied research, (3) strengthening the infrastructure for public health and providing training opportunities, and (4) developing, implementing and evaluating strategies for prevention and control⁴⁵. Activities in each of these areas are often integrated and complementary. Whatever the problem is, surveillance is necessary to detect it and to define its scope and magnitude. Applied research provides new surveillance techniques, new diagnostic methods and potential interventions. Surveillance, applied research, and prevention and control cannot be effective without appropriate infrastructure and training. Finally, all other efforts lead to prevention and control.

Table 1 Antimicrobial-resistant microbes affecting treatment and control of infectious diseases in the twenty-first century

Hospital-acquired infections
Methicillin-resistant staphylococci
Vancomycin-resistant staphylococci
Vancomycin-resistant enterococci
ESC-resistant Gram-negative bacteria
Azole-resistant <i>Candida</i>
Community-acquired infections
Multidrug-resistant pneumococci
FQ- and ESC-resistant <i>Salmonella</i> (including <i>S. typhi</i>)
Multidrug-resistant <i>Shigella</i> (including <i>Shig. dysenteriae</i>)
FQ-resistant gonococci
Multidrug-resistant <i>M. tuberculosis</i>
Drug-resistant malaria
Drug-resistant HIV

ESC, extended-spectrum cephalosporin (e.g. ceftriaxone or cefotaxime); FQ, fluoroquinolone (e.g. ciprofloxacin).

The problem of emerging infections cannot be addressed by a procrustean strategy. The relative importance of each strategy varies from problem to problem. The nature of the problem might be different in the developed and developing worlds. For example, addressing antimicrobial resistance is likely to require different strategies for different organisms and different parts of the world. In surveillance strategies, having sentinel centres in hospitals or renal dialysis units might be the most appropriate way of detecting and responding to the first emergence of vancomycin-resistant *Staph. aureus*. In contrast, for *Strep. pneumoniae*, population-based surveillance that determines the extent of resistance in different geographic areas might be better. For some infections, applied research might best be directed towards (1) new antimicrobial agents, (2) mechanisms to decrease host susceptibility through vaccines or immunomodulators, (3) new surveillance techniques, (4) genetic studies to improve our understanding of the host and pathogen, or (5) techniques to improve the use of antimicrobial agents. The infrastructures necessary to treat diseases will be different. Directly observed therapy with its extensive needs for resources is appropriate for tuberculosis but not for other infections. Although specific training needs will vary from problem to problem, a well-trained and flexible workforce will be needed as problems emerge and evolve. Prevention and control measures will also vary. Addressing antimicrobial resistance in *Salmonella* will require the prudent use of antibiotics in animal populations and strategies to prevent food-borne transmission. In contrast, the control of drug-resistant *Strep. pneumoniae* might best be addressed by the development and use of conjugate pneumococcal vaccines and the more prudent use of antimicrobial agents in humans.

History indicates that infectious diseases will remain an ever-changing problem for public health. Effectively addressing these problems will require awareness, flexibility, resources and long-term commitment so that the improvements of the twentieth century can be extended into the developing world and preserved in the developed world. □

1. Sigerist, H. E. *The Great Doctors* 372 (Dover Publications, New York, 1971).
2. Garrett, L. In: *AIDS in the World* (eds Mann, J. M., Tarantola, D. J. M. & Netter, T. W.) 825–839 (Harvard Univ. Press, Cambridge, Massachusetts, 1992).
3. Culotta, E. Funding crunch hobbles antibiotic resistance research. *Science* **264**, 362–363 (1994).
4. Bax, R. P. Antibiotic resistance: a view from the pharmaceutical industry. *Clin. Infect. Dis.* **24**(Suppl), 5151–5153 (1997).
5. Centers for Disease Control and Prevention. *1994 Fact Book 7* (CDC, Atlanta, 1994).
6. National Center for Health Statistics. *Trends and Current Status in Childhood Mortality, United States,*

- 1900–85 (Vital and Health Statistics Series 3, no. 26, US Govt Printing Office, Washington, DC, 1989).
7. McKeown, T. & Lowe, C. R. *An Introduction to Social Medicine* (F. A. Davis, Philadelphia, 1966).
8. Ackerknecht, E. H. *History and Geography of the Most Important Diseases* (Hafner, London, 1965).
9. Garrod, L. P., Lambert, H. P. & O'Grady, F. *Antibiotic and Chemotherapy 4* (Churchill Livingstone, Edinburgh, 1973).
10. Armstrong, G. L., Conn, L. A. & Pinner, Y. Trends in infectious disease mortality in the United States during the 20th century. *J. Am. Med. Assoc.* **281**, 61–66 (1999).
11. Centers for Disease Control. *Poliomyelitis Surveillance Summary 1980–81* (CDC, Atlanta, 1982).
12. National Center for Health Statistics. *Health, United States, 1999* (US Govt Printing Office, Washington, DC, 1999).
13. World Health Organization. *Removing Obstacles to Healthy Development* (World Health Organization, Geneva, 1999).
14. Institute of Medicine. *Emerging Infections: Microbial Threats to Health in the United States* (National Academy Press, Washington, DC, 1992).
15. US Department of Commerce, Bureau of the Census. *1990 Census of Population. General Population Characteristics* (US Govt Printing Office, Washington, DC, 1990).
16. Ezzell, C. Care for a dying continent. *Sci. Am.* **282**(5), 96–105 (2000).
17. Centers for Disease Control and Prevention. *National Prevalence of Diabetes in the U.S. Diabetes Fact Sheet* (CDC, Atlanta, 1995).
18. American Cancer Society. *Cancer Facts and Figures — 1996* (American Cancer Society, Atlanta, 1996).
19. Fisher-Hoch, S. P. & Hutwagner, L. Opportunistic candidiasis: an epidemic of the 1980s. *Clin. Infect. Dis.* **21**, 897–904 (1995).
20. Dixon, D. M., McNeil, M. M., Cohen, M. L., Gellin, B. G. & LaMontagne, J. R. Fungal infections: a growing threat. *Publ. Hlth Rep.* **111**, 226–239 (1996).
21. Nuorti, J. P. et al. Cigarette smoking and invasive pneumococcal disease. *N. Engl. J. Med.* **342**, 681–689 (2000).
22. Marx, R., Aral, S. O., Rolfs, R. T., Sterk, C. E. & Kahn, J. G. Crack, sex, and STD. *Sex. Transm. Dis.* **18**, 92–101 (1991).
23. Altekruse, S. F., Cohen, M. L. & Swerdlow, D. L. Emerging foodborne diseases. *Emerg. Infect. Dis.* **3**, 285–293 (1997).
24. Ampel, N. M. et al. Coccidioidomycosis in Arizona: increase in incidence from 1990 to 1995. *Clin. Infect. Dis.* **27**, 1528–1530 (1998).
25. Institute of Medicine. *Antimicrobial Resistance: Issues and Options* (National Academy Press, Washington, DC, 1998).
26. McCraig, L. F. & Hughes, J. M. Trends in antimicrobial drug prescribing among office-based physicians in the United States. *J. Am. Med. Assoc.* **273**, 214–219 (1995).
27. Hennessy, T. W. et al. A national outbreak of *Salmonella enteritidis* infections from ice cream. *N. Engl. J. Med.* **334**, 1281–1286 (1996).
28. Garrett, L. *The Coming Plague: Newly Emerging Diseases in a World out of Balance* (Farrar, Straus, & Giroux, New York, 1993).
29. Hjelle, B. & Glass, G. E. Outbreak of hantavirus infection in the Four Corners region of the United States in the wake of the 1997–1998 El Niño — Southern Oscillation. *J. Infect. Dis.* **181**, 1569–1573 (2000).
30. Daniels, N. A. et al. *Vibrio parahaemolyticus* infections in the United States, 1973–1998. *J. Infect. Dis.* **181**, 1661–1666 (2000).
31. Centers for Disease Control and Prevention. Outbreaks of influenza A infections among travelers — Alaska and the Yukon territories. *Morbidity. Mortal. Weekly Rep.* **48**, 545–556 (1999).
32. Centers for Disease Control and Prevention. Outbreak of West Nile-like viral encephalitis — New York, 1999. *Morbidity. Mortal. Weekly Rep.* **48**, 845–848 (1999).
33. Herwaldt, B. L., Ackers, M. L. and the Cyclospora Working Group. An outbreak in 1996 of cyclosporiasis associated with imported raspberries. *N. Engl. J. Med.* **336**, 1548–1556 (1997).
34. Whittam, T. S. et al. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* **61**, 1619–1629 (1993).
35. Bloom, B. R. & Murray, C. J. L. Tuberculosis: commentary on a reemerging killer. *Science* **257**, 1055–1064 (1992).
36. Cantwell, M. F. & Binkin, N. J. Tuberculosis in sub-Saharan Africa: a regional assessment of the impact of the human immunodeficiency virus and National Tuberculosis Control Program quality. *Tubercle Lung Dis.* **77**, 220–225 (1996).
37. Mead, P. S. et al. Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5**, 607–625 (1999).
38. Cohen, M. L. Epidemiology of drug resistance: implications for a post-antimicrobial era. *Science* **257**, 1050–1055 (1992).
39. Centers for Disease Control and Prevention. National nosocomial infections surveillance (NNIS) report, data summary from January 1990–May 1999, issued June 1999. *Am. J. Infect. Control* **27**, 520–532 (1999).
40. Centers for Disease Control and Prevention. *Staphylococcus aureus* with reduced susceptibility to vancomycin. *Morbidity. Mortal. Weekly Rep.* **48**, 1165–1167 (2000).
41. Ries, A. A. et al. Epidemic *Shigella dysenteriae* type 1 in Burundi: panresistance, and implications for prevention. *J. Infect. Dis.* **169**, 1035–1041 (1994).
42. Lee, L. A. et al. Increase in antimicrobial-resistant *Salmonella* infections in the United States, 1989–1990. *J. Infect. Dis.* **170**, 18–34 (1994).
43. Centers for Disease Control and Prevention. *National Antimicrobial Resistance Monitoring System: Enteric Bacteria. 1998 Annual Report* (CDC, Atlanta, 1999).
44. Centers for Disease Control and Prevention. Geographic variation in penicillin resistance in *Streptococcus pneumoniae* — selected sites, United States, 1997. *Morbidity. Mortal. Weekly Rep.* **48**, 656–661 (1999).
45. Centers for Disease Control and Prevention. *Preventing Emerging Infectious Diseases. A Strategy for the 21st Century* (US Department of Health and Human Services, Atlanta, 1998).

Pathogenic strategies of enteric bacteria

Michael S. Donnenberg

Department of Medicine, Division of Infectious Diseases, University of Maryland, Baltimore, 10 South Pine Street, Baltimore, Maryland 21201, USA

Enteric bacteria use a limited array of macromolecular systems to implement diverse pathogenic strategies. The cellular targets of several enteric virulence factors have recently been identified. The themes that have emerged from these studies include the exploitation of molecules that regulate the actin cytoskeleton and the activation of apoptotic pathways to serve the pathogen.

Enteric Gram-negative pathogens cause the bulk of the burden of diarrhoea and enteric fever that leads to more than three million deaths each year. Perhaps because of the close genetic relationships between these bacteria and the strains of *Escherichia coli* used in many of the pioneering studies in the field of bacterial genetics, investigators studying these bacteria enjoyed a head start in gene discovery that has translated into a rapidly improving understanding of the pathogenesis of infections caused by these organisms. The emergence of the fields of cellular and molecular microbiology has been fuelled in large part by contributions from investigators studying enteric pathogens. More cellular targets of bacterial virulence factors are continually being identified and an image of the cell biology of infection is being built up. Here I highlight some of the recent advances in understanding common mechanisms and specific strategies by which enteric bacteria subvert host cell processes to cause disease.

The origin of enteric bacterial virulence factors

The species *E. coli* and *Salmonella enterica* shared a common ancestor about 140 Myr ago¹. *Shigella* is also closely related, considered by most experts to belong within the genus *Escherichia*. *Yersinia* and *Vibrio* are more distantly related genera, but share many features with *E. coli* at the genomic and phenotypic levels. The genome of *E. coli*, even of the non-pathogenic K-12 strain used in many laboratories, shows evidence of tremendous plasticity. Despite overall constraints on genome size, an estimated 17.6% of its coding sequence was acquired from other species². There are at least eight pathovars of *E. coli*, each of which has a specific set of virulence attributes and causes a clinical syndrome that is more or less distinct. This tremendous pathogenic diversity superimposed on a relatively conserved genetic backbone is a result of the evolution and dispersion of virulence genes that are spread between the species on transmissible genetic elements. These elements include plasmids, bacteriophages (viruses that infect bacteria) and pathogenicity islands. Pathogenicity islands are defined as discrete and relatively large loci containing sets of virulence genes that are present within the genome of pathogenic bacteria but absent from closely related non-pathogenic strains. Their origin in other species is suggested by their nucleotide content, which usually differs from the rest of the host bacterial genome, and by the frequency of insertion sequences within or flanking the

islands. The evolutionary histories of these elements are obscured by subsequent insertions, deletions and rearrangements. Further complicating matters, it has recently been realized that several of these islands contain or comprise bacteriophage genomes or encode receptors for bacteriophages, which might themselves encode virulence factors^{3–5}.

Pathogenic mechanisms used by enteric bacteria

The traditional infectious cycle — entry of the pathogen, establishment and multiplication, avoidance of host defences, damage and exit — is executed by enteric pathogens using conserved macromolecular systems. The specifics of the bacterial effector molecules used and the resulting effects on the host vary from pathogen to pathogen, but the underlying platforms for presenting the factors that interact with the host are remarkably conserved. Two types of system that contribute to the cycle will be described here: adhesins, which allow bacterial establishment, and secretion systems, which damage the host and permit the avoidance of host defences.

Pilus and non-pilus adhesins

Enteric pathogens use a variety of highly specific adhesins to attach to receptors on the surface of enterocytes and to avoid being carried away by peristalsis. To overcome the net repulsive force caused by the negative surface charges of both bacterium and host cell, these adhesins can be poised at the tip of surface appendages known as pili or fimbriae, which protrude like hairs 1–2 μm from the bacterial surface. Alternatively, the adhesins can be anchored in the bacterial outer membrane and protrude outward.

Pili. Pili are composed of many subunits of a major structural protein (pilin) packed tightly into a helical array. In a classic example of convergent evolution, the two major types of fimbria that are important for interactions between enteric pathogens and their hosts, the chaperone–usher pilus family and the type IV pilus family, are totally unrelated to each other. But within each group the machinery for pilus assembly is highly conserved.

Several types of fimbria from the chaperone–usher family of pili are composite structures composed of a flexible fibrillar tip joined end-to-end to a rigid rod. The adhesin protein, which binds surface carbohydrates on host cells, sits at the distal end of the structure. Pilus subunits are secreted across the bacterial inner membrane via the general secretory pathway into the periplasmic space, where they are bound by a specific chaperone that caps interactive surfaces

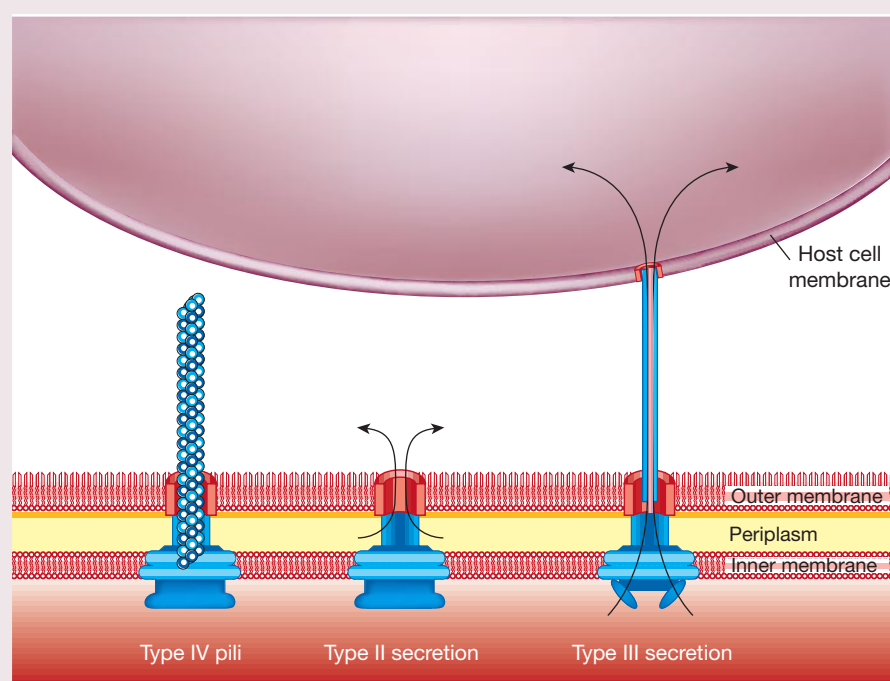


Figure 1 Macromolecular systems for type IV pilus, type II secretion and type III secretion. Many of the proteins required for the biogenesis of type IV fimbriae and for the secretion of proteins through the main terminal branch of the general secretory pathway are similar and reside in the inner membrane. Both systems, as well as type III secretion systems, use a multimeric outer-membrane protein known as the secretin, which has a central pore and forms a gated channel (red). Proteins secreted via the type II system are first exported by the general secretory apparatus (not shown) to the periplasmic space and then outside the bacterium into the surrounding medium. No periplasmic phase of export has been described for type III secretion systems. The proteins exported by way of type III systems are translocated into the host cell, presumably through a pore formed by the translocation apparatus.

to prevent premature interactions between subunits⁶. The pilus–chaperone complex is delivered to the outer-membrane usher protein, which forms a channel with a diameter sufficient to allow the passage of pilin subunits but not pili, and serves as a platform for growth of the pilus structure.

The type IV pilus system is far less well understood. At least 12 proteins are required for biogenesis of these fimbriae, even though the final organelle is usually composed of a single protein^{7,8}. Several of the proteins required for biogenesis are related to proteins required for type II secretion systems (see below). In contrast to the chaperone–usher family, distinct stages of type IV fimbrial biogenesis have not been described, suggesting that the pili might be formed in a single step by a multi-component machine composed of the proteins required for biogenesis (Fig. 1). Type IV fimbriae often aggregate into bundles and the bacteria that produce them similarly form dynamic aggregates that might be important for infection.

Non-pilus adhesins. The crystal structures of the extracellular domains of two related bacterial outer membrane proteins, invasins from *Yersinia pseudotuberculosis* and intimin from enteropathogenic *E. coli* (EPEC), have recently been solved and are remarkably similar. Both adhesins are rigid rods composed of a series of immunoglobulin-like domains capable of extending from the surface of the bacterium^{9,10}. At the end of the rod is an incomplete lectin-like domain that contains the receptor-binding site, but this domain lacks a full carbohydrate-binding fold. Instead the receptor is engaged at one side of this domain.

Secretion systems

Gram-negative bacteria regulate the transport of molecules to and from the environment through both an inner and an outer membrane separated by a periplasmic space. Many proteins, such as toxins, that are important in pathogenesis must be exported from the bacterial cytoplasm, where they are synthesized to the extracellular space. Such proteins must therefore be transported across both of these membrane barriers. At least five systems have evolved to solve this problem. These systems range from the simple, in which the exported protein has a domain that serves as an outer membrane pore through which the rest of the protein passes, to the complex, involving 20 or more proteins in an intricate apparatus. Only two of these systems are discussed here (Fig. 1).

Type II secretion systems. The type II secretion system is also known as the main terminal branch of the general secretory pathway. After

cleavage of a typical signal sequence, the general secretory apparatus secretes proteins such as cholera toxin that are exported via this pathway into the periplasmic space. The type II secretion system consists of 12–14 proteins that are required for the fully folded periplasmic protein to be exported from the periplasmic space past the outer membrane¹¹. A central paradox of type II systems is that most of these proteins reside in the inner membrane. It is not clear why so many inner-membrane proteins are required for secretion of proteins that have already passed this barrier. The secretion machinery shares several proteins with the biogenesis machinery for type IV pili and with the export apparatus for filamentous bacteriophages, indicating a conserved origin and functional similarities. These machines also share with type III secretion systems a multimeric outer-membrane protein known as the secretin that forms a gated membrane channel¹². Interestingly, the size of the channel is correlated with the size of the macromolecule that is exported.

Type III secretion systems. Many enteric pathogens possess a specialized system known as the type III secretion apparatus for delivering toxins to host cells. This system not only exports proteins through both inner and outer membranes of the bacteria, but also acts as a macromolecular syringe to inject toxins directly into target cells. Many of the 20 or more genes that encode the type III secretion apparatus are conserved between species and include several loci in common with the bacterial flagellar apparatus. Indeed, it has recently been reported that the flagellar apparatus itself can secrete virulence factors¹³. A picture of the type III secretion machine is being constructed from high-resolution electron micrographs. The supramolecular structure of a type III system from *S. enterica* serotype Typhimurium resembles the flagellar apparatus, consisting of a needle protruding from the bacterial surface connected to a base that is surrounded by two upper (outer-membrane) and two lower (inner-membrane) rings¹⁴. The structure of a system from *Shigella flexneri*, which was observed by using a different method, also reveals a protruding needle, but one that is attached to a neck with a length (10 nm) sufficient to traverse both membranes, and to a large cytoplasmic bulb¹⁵. The needle in both images seems to be much thinner than the appendage identified on the surface of EPEC. This appendage forms a bridge between the bacteria and the host cell and is composed of a protein (EspA) that is secreted by the type III apparatus¹⁶. Several type III secretion systems have also been associated with the ability to form channels in lipid bilayers and to lyse erythrocytes, providing a

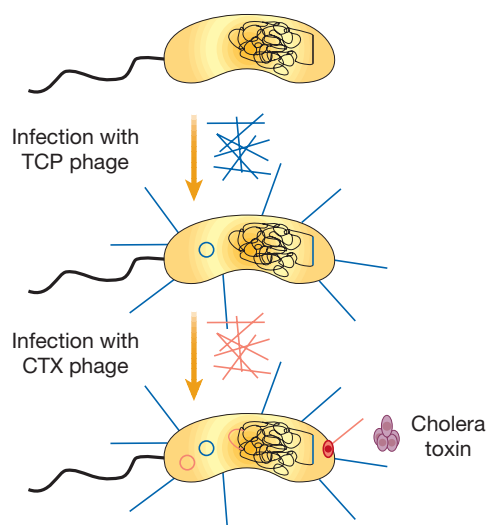


Figure 2 Acquisition of virulence factors by *V. cholerae*. Infection with TCP ϕ (blue) was a critical step in the evolution of virulence by non-pathogenic strains of *V. cholerae*. This bacteriophage/type IV fimbria serves not only as a colonization factor, but also as a receptor for the CTX ϕ encoding cholera toxin (pink). Both bacteriophages can integrate into the bacterial genome and form episomal replication intermediates (circles). The production of cholera toxin and the biogenesis of CTX ϕ both depend on a secretin (red) encoded within the bacterial genome.

plausible mechanism by which they can deliver proteins into host cells^{15,17}. The currently prevailing hypothesis states that the needle (or translocon) is able to penetrate the host membrane because of pore-forming transmembrane proteins that are exported via the apparatus (which are perhaps located at the tip of the needle). Proteins destined for the host cell travel through a putative hollow core in the translocon in a manner similar to the export of flagellin through the flagella during the growth of this structure.

Specific pathogens and infection strategies

Here I review briefly the pathogenic strategies of several enteric pathogens to illustrate how different organisms use the molecular systems described above to accomplish varied effects and cause different diseases.

Vibrio cholerae: evolution, adherence and toxin production

Vibrio cholerae is the causative agent of cholera, but only a small proportion of *V. cholerae* strains are capable of causing fulminant diarrhoea. The organism's natural habitat is estuarine and coastal waters, where it can be found living free and associated with invertebrates. Most strains lack the toxin and adhesin virulence factors necessary for causing severe disease in humans. The molecular archaeology of *V. cholerae* is a fascinating example of the role of transmissible genetic elements in the evolution of a pathogen.

Type II secretion and cholera toxin. The principal virulence factor for *V. cholerae* is cholera toxin, which induces a cascade of events beginning with NAD-ribosylation of the α -subunit of the G_s heterotrimeric receptor and culminating in the opening of the cystic fibrosis transmembrane conductance regulator. The result is a stimulation of chloride secretion by intestinal crypt cells and a blockade of sodium absorption in villous cells, leading to voluminous fluid efflux into the intestinal lumen. Export of the toxin by the bacteria is dependent on a type II secretion apparatus, which also permits the secretion of extracellular enzymes. The toxin is encoded on a small genetic locus, surrounded by repeated elements, that is absent from the chromosome of non-pathogenic strains. This element is now known to be the

incomplete genome of a filamentous bacteriophage known as CTX ϕ ³. The CTX ϕ genome lacks the gene encoding the secretin, which is required for biogenesis and is part of the genome of other filamentous bacteriophages. CTX ϕ uses instead the secretin encoded within the *V. cholerae* gene cluster specifying its type II secretion apparatus¹⁸. Thus the bacteriophage has subverted a host protein, which is used both for phage biogenesis and for the export of a phage-encoded toxin.

Type IV pili and bacteriophage. Pathogenic *V. cholerae* strains produce a type IV fimbria termed the toxin co-regulated pilus (TCP), which is absolutely essential for the colonization of humans and for virulence. Similar to other type IV pili, TCP mediates auto-aggregation of the bacteria. Furthermore, TCP is the receptor for CTX ϕ ³. Thus, strains that can colonize the human host are preferentially capable of acquiring the genes encoding cholera toxin. The genes required for TCP biogenesis are encoded on what was initially thought to be a pathogenicity island that is found in all pathogenic strains and in some non-pathogenic strains of *V. cholerae*, but it was subsequently determined that TCP is itself a filamentous bacteriophage with a specific integration site within the genome of some but not all *V. cholerae* strains⁴. Orchestrating the expression of the toxin and adhesin essential for virulence is a hierarchy of regulatory proteins, one of which is encoded within the TCP gene cluster and others elsewhere in the genome. Thus, the emergence, literally from the depths of the oceans, of this potentially fatal pathogen depended on the sequential acquisition of genetic elements including regulatory factors, a bacteriophage insertion site, a bacteriophage that serves as both a colonization factor and a receptor for a second bacteriophage, and a bacteriophage encoding a potent toxin (Fig. 2).

EPEC: adherence and cytoskeletal disruption

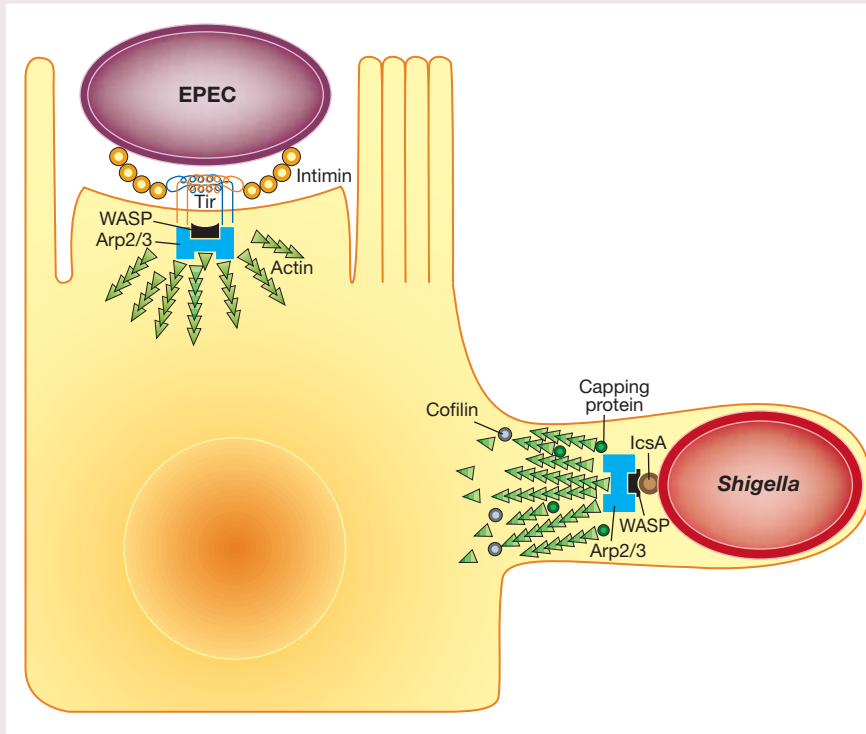
Although most *E. coli* strains live harmlessly in the mammalian gastrointestinal tract, there exist several types of pathogenic strains possessing specific virulence factors that enable them to cause diarrhoea or extraintestinal infections. Among these virulent strains, EPEC are noteworthy for both the severity of the disease that they cause and for the intricacies of their pathogenesis¹⁹. First identified in the 1940s as the cause of devastating outbreaks of neonatal gastroenteritis, EPEC are still a leading cause of diarrhoea in infants in developing countries.

Reversible aggregation and type IV pili. EPEC produce a type IV fimbria known as the bundle-forming pilus (BFP), so named because of the tendency for the individual fibres to aggregate into rope-like bundles. This bundling property results in the formation of large bacterial aggregates. Both the bacterial aggregates and the bundled fibres are dynamic. During adhesion to host cells, the bundles of pili organize into higher-order cables of increasing diameter and length²⁰; a protein known as BfpF is required for both formation of the cables and dispersal of the aggregates. In comparison to the wild-type strain, a *bfpF* mutant is hyperpiliated and hyperadherent, forms aggregates that fail to disperse, and has a decrease in virulence in volunteers comparable to that of a mutant that produces no pili at all^{21,22}. Exactly how BfpF, which is predicted to be a cytoplasmic nucleotide-binding protein, accomplishes these diverse activities is not known; neither is the role of bacterial aggregation and dispersal in virulence.

Adhesins, type III secretion, receptors and cytoskeletal reorganization. The hallmark of EPEC infection is the formation of adhesion pedestals in cells to which they attach intimately. These cells lose their microvilli and develop actin-rich protrusions on which the bacteria rest in a process known as attaching and effacing. A pathogenicity island known as the locus of enterocyte effacement, which encodes a type III secretion apparatus, is necessary and sufficient for this attaching and effacing activity. Proteins secreted via this apparatus include the EspA filament, which forms a bridge between the bacteria and host cells¹⁶, other Esp proteins with roles in pathogenesis that have not yet been defined, and Tir. Tir is the receptor for intimin, which is also encoded in the locus of enterocyte effacement. The type III system and several of the Esp secreted proteins are required for Tir to be

Figure 3 Exploitation of actin polymerization machinery by enteric pathogens. Enteropathogenic *E. coli* (EPEC) (purple) causes the localized destruction of microvilli and the formation of actin-rich pedestals, a process known as attaching and effacing. Attaching and effacing requires the delivery, by the type III secretion system, of the translocated intimin receptor (Tir, orange and dark blue), which is inserted into the host membrane. Tir is a dimer; each monomer is composed of two anti-parallel α -helices joined by a loop to form a four-helix bundle¹⁰. Two molecules of the bacterial outer-membrane adhesin intimin (yellow) contact the loops at opposite ends of the dimer in a plane roughly parallel to the surfaces of both the bacteria and the host cell, explaining the intimate adhesion characteristic of EPEC infection. EPEC recruits members of the Wiscott–Aldrich syndrome protein (WASP, black) family to areas of attachment⁴⁰. WASP is normally activated by the small GTPase Cdc42, but EPEC does not require Cdc42 to form actin pedestals, suggesting that they activate WASP in an alternative manner⁴¹. WASP in turn recruits and activates the Arp2/3 complex (blue), which nucleates actin (light green) and stimulates the formation of actin filaments. It remains to be determined whether Tir is directly involved in the recruitment of WASP and whether additional EPEC secreted proteins might be involved. *Shigella* (red) recruits some of the same proteins for intracellular motility, eventually forming pseudopodia that are engulfed

by adjacent cells to allow spreading intercellularly. IcsA (brown) on the surface of the bacteria binds to and activates N-WASP (black). The binding of IcsA greatly increases the affinity of N-WASP for the Arp2/3 complex and enhances the ability of the latter to nucleate and assemble actin filaments. N-WASP might act as a molecular ratchet, binding actin filaments at the bacterial surface and, in association with the Arp2/3 complex, adding actin monomers to one end of the filaments⁴². Movement *in vitro* depends on two additional proteins: actin-depolymerizing factor, also known as cofilin (silver), and capping protein (dark green)²⁸. The former is required to provide a sufficient steady-state level of actin monomers by accelerating their release from the distal end of filaments. The latter is required to ensure that new actin monomers are added preferentially to filaments at the bacterial surface to allow directional movement.



translocated to the host cell, where it localizes in the membrane and serves as the receptor for intimin²³. Thus, rather than searching for a receptor to which they can bind, the bacteria carry around their own receptor and insert it into the host cell. An early understanding of the cellular events that lead to pedestal formation is being developed (Fig. 3). Some of the same host proteins that are subverted for actin-based intracellular motility by other enteric bacteria are involved in the formation of the actin-rich pedestal structures by EPEC. It is therefore no surprise that these EPEC-induced cellular structures are in fact dynamic²⁴.

***Shigella*: invasion and spread, and induction of cytokines**

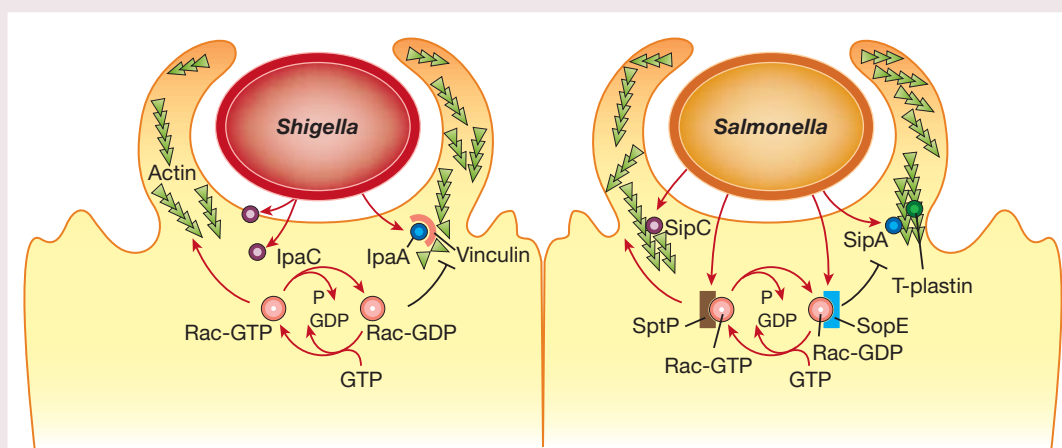
The disease bacillary dysentery, characterized by the frequent passage of mucoid, bloody stools accompanied by pain and fever, is caused by bacteria assigned to the genus *Shigella*. These infections occur in outbreaks in developed countries but are prevalent throughout the developing world, where they result in a high mortality compared with most causes of diarrhoea. The pathogenesis of *Shigella* infections depends primarily on a large plasmid that encodes factors permitting the bacteria to enter cells, grow free in the cytoplasm and spread directly to adjacent cells²⁵. The plasmid also encodes factors that stimulate the apoptosis of macrophages and the release of cytokines.

Cell entry. *Shigella* strains readily invade epithelial cells from the basolateral surface; this invasion is dependent on a type III secretion apparatus encoded on the large plasmid and on the mammalian small GTPases Cdc42, Rac and Rho, which regulate the actin cytoskeleton (Fig. 4). Two proteins secreted via the type III apparatus, IpaB and IpaC, are inserted into host cell membranes¹⁵ and seem to mediate cell entry directly, because latex beads coated with these proteins are taken up by cells²⁶.

Recruitment of actin for intracellular movement. On entering a cell, *Shigella* rapidly lyses the vacuolar membrane to become free in the cytoplasm. The bacteria become enveloped in a ‘cloud’ of actin and then begin to move through the cytoplasm, leaving a trail of actin filaments behind. Pseudopodia containing bacteria trailed by actin are formed and engulfed by neighbouring cells, allowing *Shigella* to spread from cell to cell. This actin-based intracellular motility and the ability to spread directly from cell to cell are shared with several other intracellular pathogens, including the Gram-positive bacterium *Listeria monocytogenes*, several species of *Rickettsia* and even vaccinia virus²⁷. In *Shigella* both processes depend on an outer-membrane protein called IcsA. Recently the entire motility apparatus has been reconstituted *in vitro* from purified proteins²⁸ (Fig. 3). A mutant with a null *icsA* allele is unable to spread from cell to cell and causes only small foci of infection overlying Peyer’s patches in a simian model of infection²⁵. This observation is consistent with other evidence that initial intestinal invasion occurs through specialized M cells overlying Peyer’s patches²⁹.

Apoptosis of macrophages and release of cytokines. When exposed to *Shigella*, macrophages undergo apoptosis and release large quantities of the proinflammatory cytokine interleukin (IL)-1. The IpaB protein is required for these effects, independently of its role in cell entry. Inside cells, IpaB binds to caspase-1, which is both an activator of programmed cell death and a protease that cleaves pro-IL-1 to release the mature protein (Fig. 5). Presumably, binding by IpaB activates caspase-1. The release of IL-1 in turn has a crucial role in disease because animals treated with an IL-1 antagonist have not only a diminished inflammation, but decreased intestinal invasion as well³⁰. Thus, a comprehensive model of *Shigella* infection is being built up. According to this model the bacteria breach the intestinal epithelium

Figure 4 Enteric pathogens target the cytoskeleton to induce their uptake into epithelial cells. *Shigella* (red) and *Salmonella* (orange) use similar proteins to invade epithelial cells. Both pathogens require the small actin-regulatory GTPases Rac (pink) and Cdc42 (not shown) for entry, and both enter preferentially through M cells, which are specialized epithelial cells overlying lymphocyte-rich Peyer's patches. Entry by *Shigella* requires the type III secreted protein IpaC (purple).



When IpaC is microinjected into cells, they produce actin-rich spike-like and ruffle-like surface extensions known as filopodia and lamellipodia respectively, which are similar to those seen during bacterial entry. Dominant-negative forms of Cdc42 and Rac inhibit the production of these processes⁴³. IpaA (dark blue), another protein secreted by *Shigella* by way of its type III apparatus, also has a role in cell entry. Mutants with a null *ipaA* allele invade cells with an efficiency about 10% of that of wild-type strains. IpaA co-localizes with vinculin (pink) and actin at the site of bacterial entry and binds vinculin *in vitro*⁴⁴. The IpaA–vinculin complex binds to actin and promotes the depolymerization of actin filaments (light green)⁴⁵. It is not yet known how this process augments *Shigella* invasion, however. The SipA protein (dark blue) of *Salmonella* resembles IpaA at the amino-acid level and, like *ipaA* mutants, *sipA* mutants have a reduced but residual capacity to invade cells. However, rather than binding to vinculin, SipA binds directly to actin and, indirectly through actin, to the actin-bundling protein T-plastin (dark green), which is recruited to foci of bacterial entry⁴⁶. The presence of SipA enhances the actin-bundling activity of T-plastin, but it is not known how this function might contribute to the entry process. SipC (purple) is related to the IpaC secreted effector protein of *Shigella*. Like IpaC, SipC is indispensable for invasion. Purified SipC induces actin bundling *in vitro* and, after microinjection into host cells, induces an increase in filamentous actin⁴⁷. Distinct domains of SipC have different properties, as the amino terminus of the protein is sufficient for actin-bundling activity, whereas the carboxy terminus mediates the initiation of actin filament formation. SopE (light blue) is a guanine nucleotide exchange factor that is injected via the SPI1 type III secretion apparatus into the host cell's cytoplasm. There SopE binds to the inactive (GDP-bound) forms of Cdc42 and Rac (pink), accelerating the release of GDP and the binding of GTP, which activates these signalling proteins and results in the formation of membrane ruffles⁴⁸. SptP (brown) has a motif found in GTPase-activating proteins. The bacteria inject SptP into the host cell cytoplasm through the SPI1 secretion system. SptP binds to activated Cdc42 and Rac and increases their GTPase activity, thereby stimulating them towards the GDP-bound inactive form and antagonizing the effect of SopE³⁵. Bacteria engineered to produce SptP with an altered GTPase-activating protein motif induce cells to produce membrane ruffles that persist for prolonged periods.

first through M cells. From within these cells they are able to spread directly to adjacent cells or pass through to underlying macrophages, which they kill through the induction of apoptosis while releasing IL-1. The bacteria are then free to enter enterocytes via the basolateral surface. The IL-1 attracts neutrophils, which migrate to the intestinal lumen, loosening intercellular junctions in the process and giving additional luminal organisms access to the basolateral surface for further cellular invasion²⁵.

***Yersinia* and antiphagocytosis**

There are three species of the genus *Yersinia* that are pathogenic for humans. *Y. enterocolitica* and *Y. pseudotuberculosis* are enteric pathogens capable of causing systemic disease. (*Y. pestis*, the causative agent of bubonic plague, evolved quite recently from *Y. pseudotuberculosis*, having acquired two plasmids, the ability to grow in the flea, and the ability to cause severe systemic illness³¹.) Both of the enteric *Yersinia* species pathogenic for humans can cause diarrhoea and can disseminate via the mesenteric lymph nodes to the liver, blood and spleen. They use the surface protein invasins to help breach the intestinal barrier early in the disease cycle. Within the deeper tissues, most of the organisms remain extracellular. Studies of the pathogenic mechanisms of these bacteria first led to the realization that type III secretion systems target effector molecules to host cells³². In contrast to *Shigella* and *Salmonella*, which use these systems to gain entry into non-phagocytic cells, *Y. enterocolitica* and *Y. pseudotuberculosis* use the same type of system to avoid uptake by phagocytic cells. A variety of bacterial poisons that act at different points in the regulation of the cytoskeleton are injected into host cells, where they act together to block the cytoskeletal changes required for bacterial uptake.

Type III secretion effector molecules that block bacterial uptake. The type III secretion apparatus of *Yersinia*, encoded on a large plasmid, directs the export of about 12 proteins known as Yops. Many of these Yops are

effector proteins that are translocated into host cells, where they interact with specific host proteins to disrupt cellular physiology. YopH, YopT and YopE each participate by different mechanisms in blocking bacterial entry (Fig. 5). Several additional Yops are delivered to host cells, including the serine/threonine kinase YopO (also known as YpkA), but their targets and activities are not yet known.

A type III secretion effector molecule that induces apoptosis. Like those infected with *Shigella*, macrophages infected with *Yersinia* undergo apoptosis. For both organisms, a type III secretion system is essential for this function. However, the effector proteins and signalling pathways seem to be different (Fig. 5). The protein specifically required for inducing apoptosis is known by two names, YopJ and YopP, depending on the *Yersinia* species. In a murine model of infection, a *yopJ* mutant is less virulent than the wild type, disseminates to mesenteric lymph nodes less well than the wild type and, in contrast to the wild type, fails to induce the apoptosis of macrophages *in vivo*³³. Thus, YopJ, and by inference the ability to induce apoptosis, seems to have a central role in the infection. *Yersinia* seems adept at foiling the macrophage, both by poisoning its phagocytic apparatus and by luring it to commit suicide.

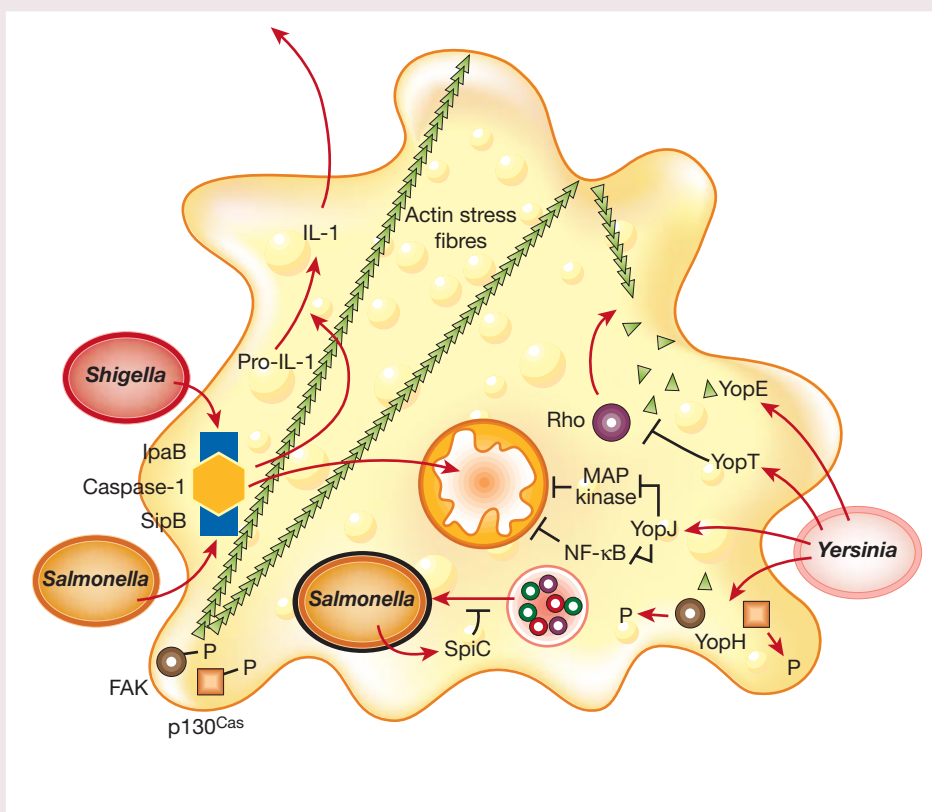
***Salmonella*: invasion, intracellular survival and systemic spread**

Salmonella enterica, a species designation that includes all of the serovars commonly associated with human illness, can cause a variety of disease syndromes, including self-limited gastroenteritis, focal infections, bacteraemia and typhoid fever. *S. enterica* owes its virulence largely to the presence in its genome of pathogenicity islands. These virulence loci encode two distinct type III secretion systems, a number of fimbriae and a complex regulatory network.

***Salmonella* pathogenicity island-I (SPI1) and cellular invasion.** The acquisition of a pathogenicity island encoding a type III secretion system that directs the uptake of the bacteria by non-phagocytic cells was a key event in the evolution of *S. enterica*. The cellular targets of several

Figure 5 Strategies used by enteric bacteria to foil macrophages. *Shigella* (red) and *Salmonella* (orange) secrete related proteins known as IpaB and SipB respectively (blue), via their type III secretion systems. Both proteins can bind to and presumably activate caspase-1 (yellow), which cleaves pro-IL-1 to its mature form and induces apoptosis. *Salmonella* also secretes a protein known as SpiC through its SPI2 secretion system; this protein blocks vesicle trafficking (depicted as a failure of fusion between lysosome and endosome). *Yersinia* (pink) delivers a variety of Yops through its type III secretion system into host cells to interfere with macrophage function. Macrophages infected with *Yersinia* produce smaller quantities of tumour necrosis factor- α , and stimulation of anti-apoptotic pathways through the activation of NF- κ B and MAP kinase is inhibited^{49,50}. The effector molecule known as YopP in *Y. enterocolitica* and YopJ in *Y. pseudotuberculosis* is targeted to the host cell cytoplasm and is required both for the inhibition of tumour necrosis factor- α production and for apoptosis. The YopH protein is a highly active tyrosine phosphatase that is directed to and disrupts focal adhesions, sites where actin stress fibres (light green) contact the cell membrane⁵¹.

Among its targets are the proteins focal adhesion kinase (FAK) and p130^{Cas} (brown and orange)^{52,53}. The result is an inhibition of bacterial uptake by epithelial cells and macrophages. YopT, another protein targeted by the type III system to the host cell's cytoplasm, participates in inducing a loss of actin stress fibres in the cells. It was recently realized that the electrophoretic mobility of Rho (purple), one of the small GTPases that regulates the actin cytoskeleton, is altered in cells infected with *Y. enterocolitica* and that this altered mobility is dependent on YopT⁵⁴. Presumably this alteration of Rho affects its function, leading to the observed cytoskeletal changes. The YopE protein is also targeted to the cytoplasm of host cells and, like YopT, disrupts the actin cytoskeleton. However, the mechanism of action of YopE is unknown.



of the effector proteins delivered to cells by this apparatus have been identified, and progress toward a comprehensive understanding of the process has been rapid. Several similarities exist between the mechanisms of cellular invasion by *S. enterica* and *Shigella*. Both organisms stimulate extensive membrane ruffling in cells during the entry process, and both require Cdc42 and Rac for entry. In addition, the SPI1 system includes homologues of the Ipa proteins of *Shigella*; these proteins seem to have similar, but not necessarily identical, functions (Fig. 4). Not all effector proteins secreted via the SPI1 type III apparatus are encoded within SPI1. The SopE protein is part of the genome of a bacteriophage that has infected strains of *S. enterica* serovars Dublin and Typhi and some strains of serovar Typhimurium³⁴. SopE directly activates Rac and Cdc42. Microinjection of cells with SopE results in membrane ruffling and can overcome the defect in invasion by bacteria that have mutations in the SPI1 secretion system.

Whereas many of the bacterial effector proteins so far identified interact with host proteins in a manner that induces the cells to change their shape, one bacterial protein that induces cells to regain their original morphology has been described. SptP acts directly on Rac and Cdc42 to antagonize the effects of SopE and return the host cell to its normal state³⁵. It seems that the bacteria are capable of rectifying some of the damage that they wrought during the entry process. Presumably this benefits both the bacteria and the host.

The SPI1 locus also encodes a homologue of IpaB, known as SipB; like IpaB, SipB is both essential for efficient invasion and activates caspase-1 to induce macrophage apoptosis (Fig. 5)³⁶.

SPI2 and intracellular survival. A second pathogenicity island in *S. enterica*, SPI2, was uncovered during a screen for mutants that competed

poorly for dissemination and survival in mice after intraperitoneal infection³⁷. In contrast to mutants with disruptions in genes within SPI1, mutants with disruptions in SPI2 invade cells normally but fail to survive and multiply within macrophages, a hallmark of *Salmonella* pathogenesis. Expression of the SPI2 secretion system is stimulated when the bacteria are inside macrophages. Phagosomes containing *Salmonella* fail to fuse with lysosomes; the phagocyte is therefore unable to deliver microbicidal compounds to the bacteria, so the bacteria replicate. Unlike normal mice, mice deficient in the ability to produce a functional NADPH oxidase, which are therefore incapable of generating a bactericidal respiratory burst, are susceptible to infection with mutants that have defects in the SPI2 type III secretion system³⁸. The NADPH oxidase does not localize to vacuoles containing wild-type *Salmonella*, but does localize to vacuoles containing bacteria with mutations within SPI2. One SPI2-encoded protein that has a role in vesicle trafficking has been identified³⁹. SpiC is secreted by the SPI2 type III secretion system and translocated into the host cell's cytosol. SpiC is capable of blocking vesicle trafficking when expressed in cells from a viral vector, and purified SpiC protein can block endosome fusion *in vitro* (Fig. 5). The target of SpiC has not yet been identified.

Conclusions

Enteric pathogens use a limited set of platforms to implement a wide range of pathogenic strategies. Remarkably, not only are the macro-molecular systems conserved, but also some of the same host cell targets are exploited, even though many of the effector molecules and the resulting diseases differ. Some of these bacteria choose to remain in the intestinal lumen, others penetrate into the intestine, and others

invade and spread systemically. Both *V. cholerae* and EPEC use a type IV pilus to form aggregates, which seem to aid colonization. The pilus from *V. cholerae* is itself both a bacteriophage and a receptor for a second bacteriophage specifying the critical cholera virulence factor. Could the EPEC type IV pilus also be a bacteriophage? EPEC, *Shigella*, *Yersinia* and *Salmonella* all use type III secretion systems to place effector proteins within host cells. EPEC uses its type III secretion system to target the WASP-Arp2/3 complex and form actin pedestals. The IcsA outer membrane protein of *Shigella* activates the same complex, but uses it to generate actin-based intracellular movement. How do these processes differ? *Shigella* and *Salmonella* use similar type III secretion effector molecules to influence the polymerization and bundling of actin filaments and gain entry into host cells. For both bacteria, small GTPase are central to these processes. Do they activate these GTPases in the same manner? The fact that not all strains of *S. enterica* serotype Typhimurium are infected with the SopE-encoding bacteriophage raises the question: what causes ruffling in *sopE*-negative strains of *Salmonella*? *Yersinia* uses its type III secretion apparatus to avoid being engulfed by phagocytes, whereas *Salmonella* uses one of its type III secretion systems to thrive within macrophages. Both of these pathogens, as well as *Shigella*, induce apoptosis in macrophages, but *Yersinia* seems to target a different pathway toward this end. How does the specific apoptotic pathway influence the infection? Investigators studying enteric pathogens have generated an abundant array of fascinating findings and an even greater collection of compelling questions. □

- Whittam, T. S. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 2708–2720 (ASM Press, Washington, DC, 1996).
- Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417 (1998).
- Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
- Karaolis, D. K., Somara, S., Maneval, D. R. Jr, Johnson, J. A. & Kaper, J. B. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379 (1999).
- Perna, N. T. *et al.* Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66**, 3810–3817 (1998).
- Thanassi, D. G., Saulino, E. T. & Hultgren, S. J. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr. Opin. Microbiol.* **1**, 223–231 (1998).
- Yoshida, T., Kim, S. R. & Komano, T. Twelve *pil* genes are required for biogenesis of the R64 thin pilus. *J. Bacteriol.* **181**, 2038–2043 (1999).
- Anantha, R. P., Stone, K. D. & Donnenberg, M. S. Effects of *bfp* mutations on biogenesis of functional enteropathogenic *Escherichia coli* type IV pili. *J. Bacteriol.* **182**, 2498–2506 (2000).
- Hamburger, Z. A., Brown, M. S., Isberg, R. R. & Bjorkman, P. J. Crystal structure of invasins: a bacterial integrin-binding protein. *Science* **286**, 291–295 (1999).
- Luo, Y. *et al.* Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex. *Nature* **405**, 1073–1077 (2000).
- Russel, M. Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *J. Mol. Biol.* **279**, 485–499 (1998).
- Marciano, D. K., Russel, M. & Simon, S. M. An aqueous channel for filamentous phage export. *Science* **284**, 1516–1519 (1999).
- Young, G. M., Schmiel, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: the flagellar export apparatus functions as a protein-secretion system. *Proc. Natl Acad. Sci. USA* **96**, 6456–6461 (1999).
- Kubori, T. *et al.* Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602–605 (1998).
- Blocker, A. *et al.* The tripartite type III secretin of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J. Cell Biol.* **147**, 683–693 (1999).
- Knutton, S. *et al.* A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J.* **17**, 2166–2176 (1998).
- Tardy, F. *et al.* *Yersinia enterocolitica* type III secretion-translocation system: channel formation by secreted Yops. *EMBO J.* **18**, 6793–6799 (1999).
- Davis, B. M. *et al.* Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXφ. *Science* **288**, 333–335 (2000).
- Donnenberg, M. S., Kaper, J. B. & Finlay, B. B. Interactions between enteropathogenic *Escherichia coli* and host epithelial cells. *Trends Microbiol.* **5**, 109–114 (1997).
- Knutton, S., Shaw, R. K., Anantha, R. P., Donnenberg, M. S. & Zorngani, A. A. The type IV bundle-forming pilus of enteropathogenic *Escherichia coli* undergoes dramatic alterations in structure associated with bacterial adherence, aggregation and dispersal. *Mol. Microbiol.* **33**, 499–509 (1999).
- Anantha, R. P., Stone, K. D. & Donnenberg, M. S. The role of BfpF, a member of the PilT family of putative nucleotide-binding proteins, in type IV pilus biogenesis and in interactions between

- enteropathogenic *Escherichia coli* and host cells. *Infect. Immun.* **66**, 122–131 (1998).
- Bieber, D. *et al.* Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science* **280**, 2114–2118 (1998).
- Kenny, B. *et al.* Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**, 511–520 (1997).
- Sanger, J. M., Chang, R., Ashton, F., Kaper, J. B. & Sanger, J. W. Novel form of actin-based motility transports bacteria on the surface of infected cells. *Cell Motil. Cytoskeleton* **34**, 279–287 (1996).
- Sansonetti, P. J., Tran, V. N. & Egile, C. Rupture of the intestinal epithelial barrier and mucosal invasion by *Shigella flexneri*. *Clin. Infect. Dis.* **28**, 466–475 (1999).
- Ménard, R., Prévost, M. C., Gounon, P., Sansonetti, P. & Dehio, C. The secreted Ipa complex of *Shigella flexneri* promotes entry into mammalian cells. *Proc. Natl Acad. Sci. USA* **93**, 1254–1258 (1996).
- Cudmore, S., Cossart, P., Griffiths, G. & Way, M. Actin-based motility of vaccinia virus. *Nature* **378**, 636–638 (1995).
- Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).
- Jensen, V. B., Harty, J. T. & Jones, B. D. Interactions of the invasive pathogens *Salmonella typhimurium*, *Listeria monocytogenes*, and *Shigella flexneri* with M cells and murine Peyer's patches. *Infect. Immun.* **66**, 3758–3766 (1998).
- Sansonetti, P. J., Arondel, J., Cavaillon, J.-M. & Huerre, M. Role of interleukin-1 in the pathogenesis of experimental shigellosis. *J. Clin. Invest.* **96**, 884–892 (1995).
- Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* **96**, 14043–14048 (1999).
- Cornelis, G. R. & Wolf-Watz, H. The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol. Microbiol.* **23**, 861–867 (1997).
- Monack, D. M., Mecas, J., Bouley, D. & Falkow, S. *Yersinia*-induced apoptosis *in vivo* aids in the establishment of a systemic infection of mice. *J. Exp. Med.* **188**, 2127–2137 (1998).
- Mirold, S. *et al.* Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *Proc. Natl Acad. Sci. USA* **96**, 9845–9850 (1999).
- Fu, Y. X. & Galán, J. E. A *Salmonella* protein antagonizes Rac-1 and Cdc42 to mediate host-cell recovery after bacterial invasion. *Nature* **401**, 293–297 (1999).
- Hersh, D. *et al.* The *Salmonella* invasin SipB induces macrophage apoptosis by binding to caspase-1. *Proc. Natl Acad. Sci. USA* **96**, 2396–2401 (1999).
- Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400–403 (1995).
- Vazquez-Torres, A. *et al.* *Salmonella* pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. *Science* **287**, 1655–1658 (2000).
- Uchiya, K. *et al.* A *Salmonella* virulence protein that inhibits cellular trafficking. *EMBO J.* **18**, 3924–3933 (1999).
- Kalman, D. *et al.* Enteropathogenic *E. coli* acts through WASP and Arp2/3 complex to form actin pedestals. *Nature Cell Biol.* **1**, 389–391 (1999).
- Ben-Ami, G. *et al.* Agents that inhibit Rho, Rac, and Cdc42 do not block formation of actin pedestals in HeLa cells infected with enteropathogenic *Escherichia coli*. *Infect. Immun.* **66**, 1755–1758 (1998).
- Egile, C. *et al.* Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J. Cell Biol.* **146**, 1319–1332 (1999).
- Van Nhieu, G. T., Caron, E., Hall, A. & Sansonetti, P. J. IpaC induces actin polymerization and filopodia formation during *Shigella* entry into epithelial cells. *EMBO J.* **18**, 3249–3262 (1999).
- Tran Van Nhieu, G., Ben-Ze'ev, A. & Sansonetti, P. J. Modulation of bacterial entry into epithelial cells by association between vinculin and the *Shigella* IpaA invasin. *EMBO J.* **16**, 2717–2729 (1997).
- Bourdet-Sicard, R. *et al.* Binding of the *Shigella* protein IpaA to vinculin induces F-actin depolymerization. *EMBO J.* **18**, 5853–5862 (1999).
- Zhou, D. G., Mooseker, M. S. & Galán, J. E. An invasion-associated *Salmonella* protein modulates the actin-bundling activity of plastin. *Proc. Natl Acad. Sci. USA* **96**, 10176–10181 (1999).
- Hayward, R. D. & Koronakis, V. Direct nucleation and bundling of actin by the SipC protein of invasive *Salmonella*. *EMBO J.* **18**, 4926–4934 (1999).
- Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galán, J. E. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815–826 (1998).
- Palmer, L. E., Hobbie, S., Galán, J. E. & Bliska, J. B. YopJ of *Yersinia pseudotuberculosis* is required for the inhibition of macrophage TNF-α production and downregulation of the MAP kinases p38 and JNK. *Mol. Microbiol.* **27**, 953–965 (1998).
- Schesser, K. *et al.* The *yopJ* locus is required for *Yersinia*-mediated inhibition of NF-κB activation and cytokine expression: The *yopJ* contains a eukaryotic SH2-like domain that is essential for its repressive activity. *Mol. Microbiol.* **28**, 1067–1079 (1998).
- Persson, C. *et al.* Localization of the *Yersinia* PTase to focal complexes is an important virulence mechanism. *Mol. Microbiol.* **33**, 828–838 (1999).
- Black, D. S. & Bliska, J. B. Identification of p130^{cas} as a substrate of *Yersinia* YopH (Yop51), a bacterial protein tyrosine phosphatase that translocates into mammalian cells and targets focal adhesions. *EMBO J.* **16**, 2730–2744 (1997).
- Persson, C., Carballera, N., Wolf-Watz, H. & Fällman, M. The PTase YopH inhibits uptake of *Yersinia*, tyrosine phosphorylation of p130^{cas} and FAK, and the associated accumulation of these proteins in peripheral focal adhesions. *EMBO J.* **16**, 2307–2318 (1997).
- Zumbihl, R. *et al.* The cytotoxin YopT of *Yersinia enterocolitica* induces modification and cellular redistribution of the small GTP-binding protein RhoA. *J. Biol. Chem.* **274**, 29289–29293 (1999).

Acknowledgements

I thank J. B. Kaper for critically reviewing the manuscript. The author's work is supported by Public Health Service awards from the National Institutes of Health.

Molecular mechanisms that confer antibacterial drug resistance

Christopher Walsh

Biological Chemistry and Molecular Pharmacology Department, Harvard Medical School, Boston, Massachusetts 02115, USA

Antibiotics — compounds that are literally ‘against life’ — are typically antibacterial drugs, interfering with some structure or process that is essential to bacterial growth or survival without harm to the eukaryotic host harbouring the infecting bacteria. We live in an era when antibiotic resistance has spread at an alarming rate^{1–4} and when dire predictions concerning the lack of effective antibacterial drugs occur with increasing frequency. In this context it is apposite to ask a few simple questions about these life-saving molecules. What are antibiotics? Where do they come from? How do they work? Why do they stop being effective? How do we find new antibiotics? And can we slow down the development of antibiotic-resistant superbugs?

Antibiotics can kill bacteria (bacteriocidal) or sometimes just nullify growth (bacteriostatic). Most antibiotics in human use as antibacterials are natural products, elaborated by one species of microbe (bacteria or fungi) as chemical weapons, often in times of crowding, to kill off other microbes in the neighbouring microenvironment. Over the past 60–70 years most antibiotics have been discovered by screening of soil samples for such natural products that kill bacteria, including known pathogens, first on culture plates and then in animal infections. These include penicillins and cephalosporins from fungi and a host of antibiotics from different strains of the filamentous bacterium *Streptomyces*, such as streptomycin, erythromycin, tetracycline and vancomycin. Semisynthetic modifications have produced second- and third-generation β -lactams of both the penicillin and cephalosporin classes whereas total synthesis has created the second-generation erythromycins — clarithromycin and azithromycin. As of the end of 1999, only the fluoroquinolones (for example, ciprofloxacin) represent a totally synthetic, significant class of antibiotic.

Targets for the main classes of antibacterial drugs

To understand how antibiotics work and, concomitantly, why they stop being effective requires a brief look at the targets for the main classes of these antibacterial drugs. As summarized in Box 1, there are three proven targets for the main antibacterial drugs: (1) bacterial cell-wall biosynthesis; (2) bacterial protein synthesis; and (3) bacterial DNA replication and repair.

Cell-wall biosynthesis

The layer of the bacterial cell wall that confers strength is the peptidoglycan, a meshwork of strands of peptide and glycan that can be covalently crosslinked (Fig. 1a). The larger the fraction of adjacent peptide strands that are connected in amide linkage by action of a family of transpeptidases, the higher the mechanical strength to osmotic lysis. Transglycosylases act on the glycan strands to extend the sugar chains by incorporation of new peptidoglycan units from *N*-acetylglucosamine- β -1,4-*N*-acetylmuramyl-pentapeptide-pyrophosphoryl-undecaprenol (lipid II). Bifunctional enzymes containing both transpeptidase and transglycosylase domains are the target sites for the killing of bacteria by the β -lactam-containing penicillins and

cephalosporins, which act as pseudosubstrates and acylate the active sites of the transpeptidases (also termed penicillin-binding proteins or PBPs)⁵ (Fig. 1b). The ring-opened, penicilloylated transpeptidases deacylate very slowly, and so occupy the enzyme active sites, preventing normal crosslinking of peptide chains in the peptidoglycan layer and leaving it mechanically weak and susceptible to lysis on changes in osmotic pressure.

In addition to penicillins and cephalosporins, the vancomycin family of glycopeptide antibiotics also target the peptidoglycan layer in the cell-wall assembly. But rather than targeting the enzymes involved in peptide crosslinking, vancomycin ties up the peptide substrate⁶ and thereby prevents it from reacting with either the transpeptidases or the transglycosylases. The net effect is the same: failure to make peptidoglycan crosslinks leads to a weaker wall that predisposes the treated bacteria to a killing lysis of the cell-wall layer. The cup-shaped undersurface of the vancomycin antibiotic makes five hydrogen bonds to the D-Ala-D-Ala dipeptide terminus of each uncrosslinked peptidoglycan pentapeptide side chain (Fig. 1c), which accounts for the high affinity of the antibiotic for its target, both in partially crosslinked walls and in the lipid II intermediate. Because β -lactams and vancomycin work on adjacent steps — substrate and enzyme — they show synergy when used in combination.

Protein synthesis

The RNA and protein machinery of the prokaryotic ribosomes is sufficiently distinct from the analogous eukaryotic machinery that there are many inhibitors of protein synthesis, targeting different steps in ribosome action, with selective antibacterial action. These include such important antibiotics as the macrolides of the erythromycin class⁷, the tetracyclines⁸ (which are products of the aromatic polyketide biosynthetic pathways) and the aminoglycosides⁹ (of which streptomycin was the founding member, supplanted now by later synthetic variants such as kanamycin) (Fig. 2a). Given the large number of molecular steps involved in initiation, elongation and termination of protein assembly by the ribosome, it is not surprising that there would be many steps of binding or catalysis that could be interdicted by these and many other classes of protein-synthesis inhibitors. This multiplicity also indicates that protein synthesis will provide a multifaceted target for new antibiotics and this is the mechanism for the action of oxazolidinones¹⁰, one of

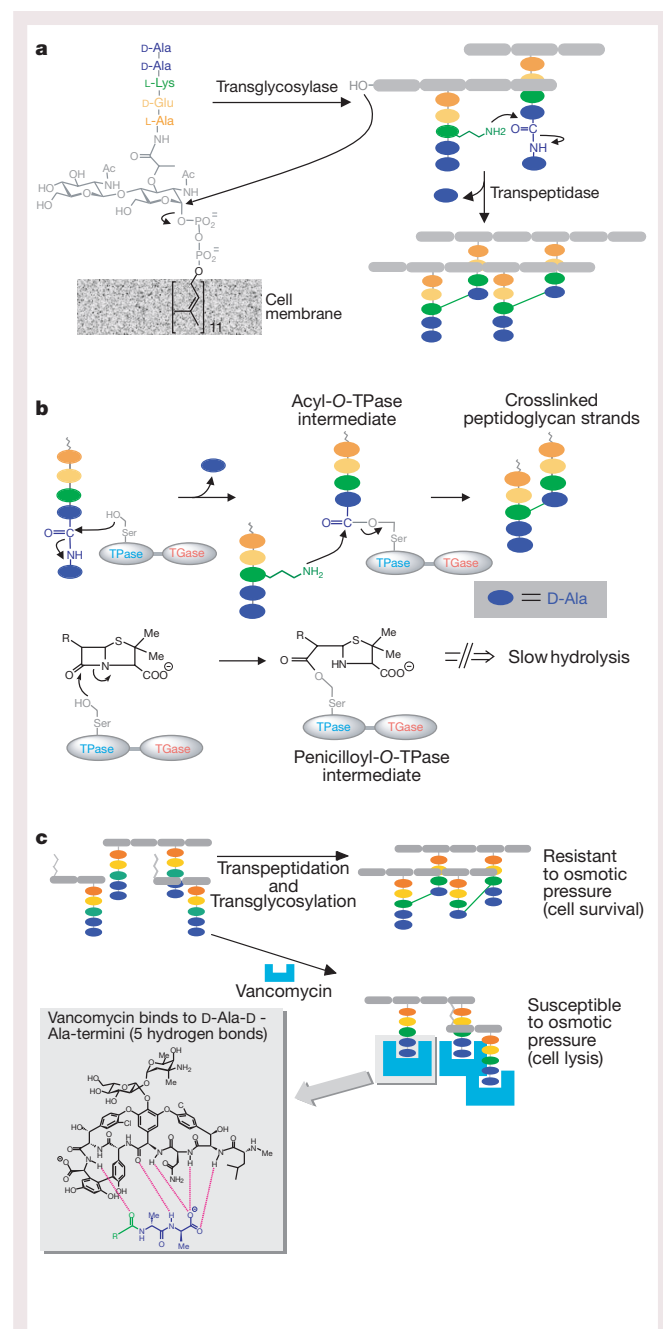


Figure 1 Blockade of transpeptidation and transglycosylation steps of cell-wall biosynthesis by penicillins and vancomycins. **a**, Interruption of the normal crosslinking and strength-conferring enzymes by antibiotics that inhibit the enzyme (penicillins) or sequester the substrate (vancomycin). **b**, Inhibition of transpeptidase activity by penicillins through formation of a slowly hydrolysing covalent acyl enzyme intermediate. **c**, Complexation of the D-Ala-D-Ala termini of peptidoglycans by vancomycin in a network of five hydrogen bonds.

which has been approved in the United States in the first quarter of 2000.

DNA replication and repair

The fluoroquinolones, such as ciprofloxacin (Fig. 2b), are synthetic antibiotic structures that kill bacteria by targeting the enzyme DNA gyrase¹¹ (Box 1), the enzyme responsible for uncoiling the inter-twined circles of double-stranded bacterial DNA that arise after each round of DNA replication. DNA topoisomerases are classified as type I or type II according to whether transient single-strand breaks (type I) or transient double-strand breaks (type II) are made in the DNA

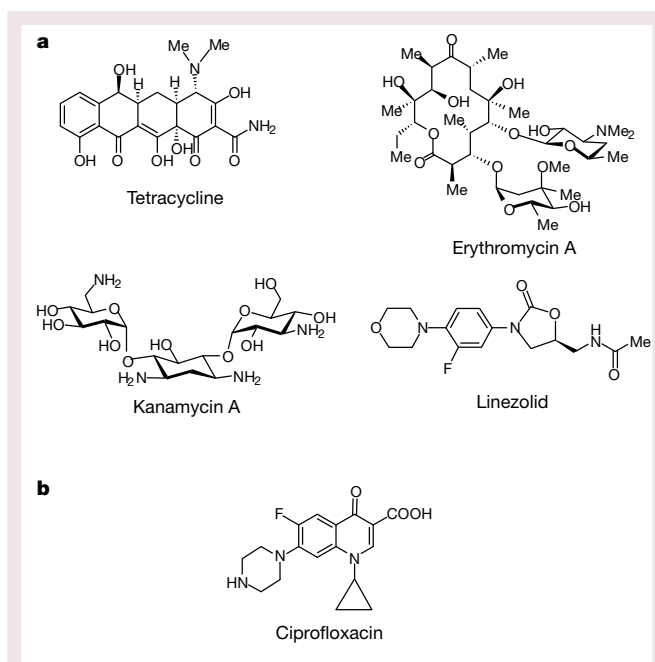


Figure 2 Structural and functional diversity of antibacterial drugs. **a**, Antibiotics comprise a diverse set of natural product structures, represented by the erythromycin class of macrolides, tetracyclines, the aminoglycosides (represented by kanamycin), and the oxazolidinone linezolid. These products target the 23S rRNA and associated proteins in the peptidyl transferase centre of the ribosome to inhibit steps in the elongation of the protein chain. **b**, The fluoroquinolones, represented by ciprofloxacin, kill bacteria by inhibiting DNA gyrase and the related topoisomerase IV in mid-catalytic cycle, by trapping a doubly cleaved DNA intermediate.

substrate to pass the DNA double helical strands through each other and reduce the linking number (the number of superhelical twists in DNA). Bacterial DNA gyrases are type II topoisomerases and the transient cleavage of both DNA strands involves the reversible attachment of the 5' ends of the cleaved DNA to tyrosyl residues on each of the two GyrA subunits in the active (GyrA)₂(GyrB)₂ tetramer¹². Quinolone antibiotics such as ciprofloxacin are mechanism-based inhibitors of DNA gyrase and act by forming a complex with the enzyme and the doubly cleaved DNA that is covalently tethered to the GyrA subunits¹¹. In the ciprofloxacin complex, the gyrase cannot religate the cleaved DNA and, as a consequence, double-strand breaks accumulate and ultimately set off the SOS repair system that leads to bacterial cell death. A second type II topoisomerase, known as topoisomerase IV, is also an important target and probably the primary one in *Staphylococcus aureus* infections¹³.

In each of the three main targets — cell wall, and protein and DNA biosynthesis — the antibiotics use comparative biochemical differences between prokaryotic machinery and eukaryotic machinery to act selectively. New classes of antibiotics that may work on additional and new targets will have to display equivalent therapeutic indices and efficacy-to-toxicity ratios to gain regulatory approval and widespread acceptance.

Bacterial survival strategies to combat antibiotics

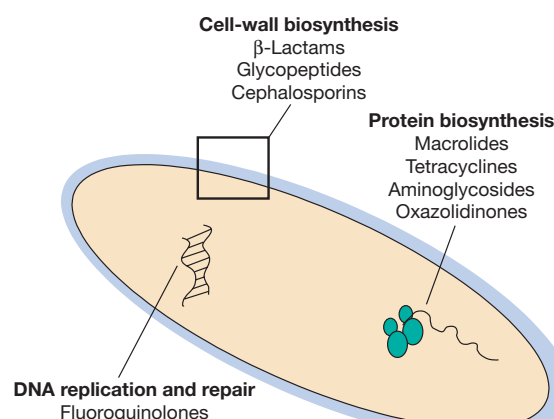
Once an antibiotic is proven to be effective and enters widespread human therapeutic use, its days are numbered. Clinically significant resistance appears in periods of months to years¹⁴. For penicillin, resistance began to be noted within two years of its introduction in the mid 1940s, which is typical for a resistance mechanism involving the action of one gene product and its time of spread through bacterial populations. Vancomycin resistance, in the context of surging sources of life-threatening vancomycin-resistant enterococci (VRE)

Box 1

Targets, mode of action and mechanisms of resistance of the main classes of antibacterial drugs

Antibiotic	Target	Mode of action	Resistance mechanism
Cell wall			
β -Lactams	Transpeptidases/transglycosylases (PBPs)	Blockade of crosslinking enzymes in peptidoglycan layer of cell walls	β -Lactamases, PBP mutants
Vancomycin	D-Ala-D-Ala termini of peptidoglycan and of lipid II	Sequestration of substrate required for crosslinking	Reprogramming of D-Ala-D-Ala to D-Ala-D-Lac or D-Ala-D-Ser
Protein synthesis			
Macrolides of the erythromycin class	Peptidyl transferase, centre of the ribosome	Blockade of protein synthesis	rRNA methylation, drug efflux
Tetracyclines	Peptidyl transferase	Blockade of protein synthesis	Drug efflux
Aminoglycosides	Peptidyl transferase	Blockade of protein synthesis	Enzymatic modification of drug
Oxazolidinones	Peptidyl transferase	Blockade of protein synthesis	Unknown
DNA replication/repair			
Fluoroquinolones	DNA gyrase	Blockade of DNA replication	Gyrase mutations to drug resistance

Box 1 Figure Proven targets for antibacterial drugs. Cell-wall biosynthesis at the stage of crosslinking of peptidoglycan peptide strands by transpeptidases and transglycosylases is inhibited by the β -lactam antibiotics (penicillins and cephalosporins). Protein biosynthesis at the ribosome is targeted by several classes of antibiotics, including macrolides, tetracyclines, aminoglycosides and oxazolidinones, which block one or more steps involving rRNA and the proteins of the ribosome at the peptidyl transferase centre. The fluoroquinolone antibiotics interrupt DNA replication by trapping a complex of DNA bound to the enzyme DNA Gyrase, a type II topoisomerase.



in hospital wards, became noticeable in 1987 and spread dramatically in the ensuing 4–6 years¹⁵. VRE have collected five genes that are necessary and sufficient for high-level resistance^{16,17} (see below) and this may explain the 29 years' delay between introduction (in 1958) and clinically important resistance.

Development of resistance is not a matter of if but only a matter of when^{3,18}. Given the large number of bacteria in an infection cycle, the rapid generation time, and the intrinsic rate of mutation of about 1 in 10^7 , then a pool of 10^{10} bacteria would have mutations on average in a thousand loci. If one of those mutations confers resistance to an applied antibiotic, whereas all sensitive bacteria are killed, the resistant one will grow, fill the space vacated by its dead neighbours and become the dominant variant in the population. If an antibiotic treatment is at subtherapeutic levels, outgrowth of resistant bacteria is practically guaranteed.

A principal mechanism for the rapid spread of antibiotic-resistance genes through bacterial populations is that such genes get collected on plasmids¹⁸ that are independently replicated within and passed between bacterial cells and species. Furthermore, some of these genes that reside on plasmids, such as the five genes that specify VRE, may be further segregated within transposons¹⁶ that can actively cut themselves out of one DNA locale and hop into other locales, promiscuously moving their antibiotic resistance-conferring genetic cargo.

Selective pressures on *S. aureus* and on *Enterococcus faecalis*, pathogens that can invade the abdominal cavity after surgical procedures, have been documented in hospital environments where a *S. aureus* culture, in a minority of patients, can switch from methicillin-susceptible *S. aureus* (MSSA) to methicillin-resistant *S. aureus* (MRSA) in five to seven days. If vancomycin is then given on day eight and cultures are sampled 14 days later, patients with signs of continuing infection are now dominated by VRE¹⁹. In addition to

showing how antibiotic-resistant bacteria are selected in hospital environments by constant antibiotic pressure, this emphasizes the continuing need for cycles of new antibiotic discovery and development.

Three types of antibiotic-resistance strategies, which deal with most of the main classes of antibacterial drugs listed in Fig. 1 and Box 1, illustrate the strategies by which resistant bacteria strike back and nullify the action of antibiotics that appear in their neighbourhoods²⁰.

Pump out the antibiotic

For antibiotics to be effective they must reach their specific bacterial targets and accumulate at concentrations that can act in some reasonable time frame. For example, the protein-synthesis machinery is located in the cytoplasm so antibacterials that are inhibitors of protein synthesis must pass through the cell membranes (outer and inner permeability barriers for Gram-negative bacteria; inner membrane barriers for Gram-positive bacteria) and then accumulate to a high enough concentration to block the particular susceptibility step of protein assembly. Both Gram-positive and Gram-negative bacteria that become resistant to tetracyclines commonly overproduce related membrane proteins (with relative molecular masses of 42,000) that act as an export or efflux pump for the drug^{21,22}. As schematized in Fig. 3a, the drug is pumped out faster than it can diffuse in, so intrabacterial concentrations are kept low and ineffectual; bacterial protein synthesis proceeds at largely unimpeded rates. The pumps are variants of membrane pumps possessed by all bacteria to move lipophilic or amphipathic molecules in and out of the cells. Some are used by antibiotic producers to pump antibiotics out of the cells as fast as they are made and so constitute an immunity or protective mechanism for the bacteria to prevent being killed by their own chemical weapons. Equivalent drug efflux pumps have been observed in several bacteria, including staphylococci, which become

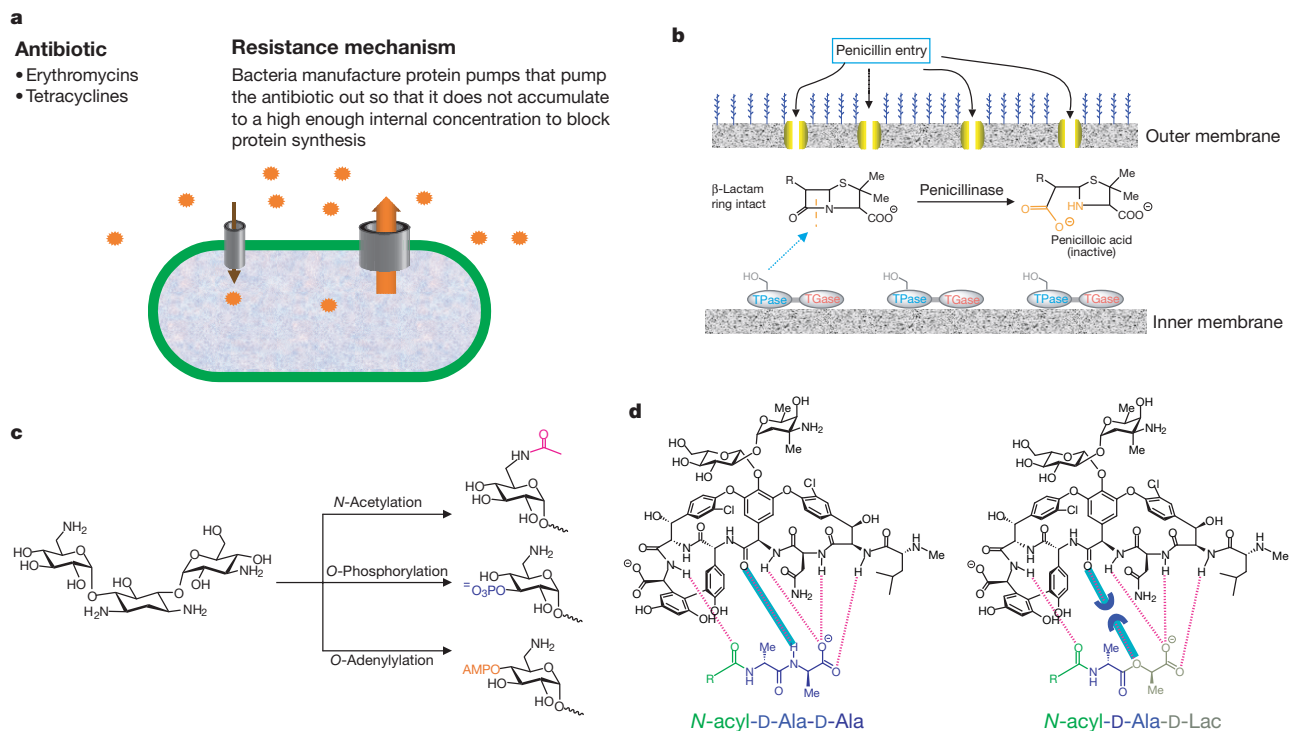


Figure 3 Principal resistance strategies for bacterial survival. **a**, Drugs such as tetracyclines or erythromycins are pumped back out of bacterial cells through efflux pump proteins to keep intracellular drug concentrations below therapeutic level. **b**, The antibiotic is destroyed by chemical modification by an enzyme that is elaborated by the resistant bacteria. This is exemplified here by the β -lactamase secreted into the periplasmic space to hydrolyse penicillin molecules before they reach their PBP targets in the cytoplasmic membrane of this Gram-negative bacterium. **c**, The aminoglycoside antibiotic kanamycin can be enzymatically modified at three sites by three kinds of enzymatic processing — *N*-acetylation, *O*-phosphorylation or *O*-adenylation — to block recognition by its target on the ribosome. **d**, The target structure in the bacterium can be reprogrammed to have a low affinity for antibiotic recognition. Here the switch from the amide linkage in the D-Ala-D-Ala peptidoglycan termini to the ester linkage in the D-Ala-D-Lac termini is accompanied by a 1,000-fold drop in drug-binding affinity.

resistant to the erythromycin class of macrolide antibiotics^{22,23}.

Destroy the antibiotic warhead

Although the above mechanism prevents the antibiotic from accumulating in the desired compartment, it leaves the antibiotic unchanged; a second strategy of resistance is destruction of the chemical warhead in the antibiotic. The classic case is the hydrolytic deactivation of the β -lactam ring in the penicillins and cephalosporins by elaboration of the hydrolytic enzyme β -lactamase by resistant bacteria²⁴ (Fig. 3b; and see discussion in refs 1, 20). Because the four-membered, strained lactam ring is the chemically activated functionality in the drugs that acylate and irreversibly modify the cell wall-crosslinking PBPs, the hydrolysed, ring-opened penicilloic acid product is now deactivated and nonfunctional as a PBP pseudosubstrate and useless as an antibiotic. The lactamase-producing bacteria secrete this enzymatic weapon into the periplasm to destroy β -lactam antibiotics before they can reach the PBP targets in the cytoplasmic membrane. A single β -lactamase molecule can hydrolyse 10^3 penicillin molecules per second. So if 10^5 enzymes are secreted per resistant cell, then 100 million molecules of penicillin are destroyed every second, which is clearly an effective strategy.

Other antibiotic classes, such as the aminoglycosides, do not contain such hydrolytically labile groups. These protein-synthesis inhibitors are still neutralized by deactivating enzymes but now the enzymes decorate the periphery of the aminoglycosides with three types of chemical substituents²⁵ that interrupt the binding to the RNA targets in the ribosome. As shown in Fig. 3c, aminoglycoside-resistance enzymes can be adenylyl transferases, which add AMP moieties, phosphoryl transferases, which add PO_3 groups, or acetyl transferases, which acetylate the amino groups of the antibiotic. The modified aminoglycoside products have considerably lower affinity for RNA and so do not bind and interrupt protein synthesis.

The X-ray structure of an antibiotic phosphotransferase indicates an evolutionary relationship to a protein kinase²⁶, defining a route by which bacteria may have recruited an enzyme for the resistance brigade.

Reprogramme the target structure

A third resistance strategy focuses not on removal or destruction of the antibiotic but on a reprogramming or camouflaging of the target in the now resistant bacteria. In the erythromycin-resistance manifold, in addition to efflux pumps, resistant bacteria have emerged that have learned to mono- or dimethylate a specific adenine residue, A2058, in the peptidyl transferase loop of the 23S RNA component of the ribosome. This modification is carried out by a methyl transferase enzyme Erm²⁷ that does not impair protein biosynthesis but does lower the affinity of all the members of the erythromycin class of drugs for the RNA, as well as for the pristinamycin class described below. The Erm mechanism is the main resistance route in drug-resistant clinical isolates of *S. aureus* and is present in erythromycin-producing organisms as a self-immunity mechanism.

An additional example of the reprogramming strategy is used by VRE to escape from vancomycin. In VRE the *vanHAX* genes encode a new pathway of enzymes that reduces pyruvate to D-lactate (*vanH*), adds D-alanine and D-lactate together to produce D-Ala-D-Lac (*vanA*), and then hydrolyses the normal metabolite D-Ala-D-Ala while sparing D-Ala-D-Lac (*vanX*)¹⁷. In this cell, only the D-Ala-D-Lac accumulates and serves as a substrate to be elongated and presented at the termini of the peptidoglycan strands (Fig. 3d). The reprogramming of peptidoglycan to end in D-Ala-D-Lac rather than the normal D-Ala-D-Ala has no effect on the crosslinking efficiency carried out by the transpeptidating PBPs, but the switch from the D,D-dipeptide terminus to D,D-depsipeptide terminus lowers the binding affinity of vancomycin by 1,000-fold²⁸ and enables the VRE to grow at

1,000-fold-higher levels of antibiotic. Finally, penicillin resistance can arise not only by β -lactamase expression, but also by mutation of penicillin-binding proteins to lower-affinity forms as well as by expression of new PBPs with low affinity for antibiotic. The acquisition by *S. aureus* of the *mecA* gene that encodes a PBP2' protein with low affinity for all β -lactam antibiotics provides the molecular basis for the MRSA phenotype^{29,30,34} that is now widely disseminated.

Development of new antibiotics that circumvent resistance

Targeting the resistance mechanisms

Given the inevitability that resistant strains of bacteria will emerge in response to widespread use of a particular antibiotic and limit its lifetime, knowledge of the principal and specific resistance mechanisms, as described above, can provide insights into strategies for development of new therapeutics. Historically, when resistance to a β -lactam antibiotic arose, medicinal chemists tinkered with the periphery of the β -lactam warhead to obtain a variant that was effective, for a time. When β -lactamase-producing strains became a significant clinical menace, attention switched to approaches to neutralize the antibiotic-destroying hydrolase, both by screening against lactamase producers and by mechanism-based inhibition of the active-site serine hydrolases. Both strategies worked. Clavulanate, a natural product from a streptomycete, was not an effective antibiotic by itself but was a suicide substrate for the lactamase³¹. In combination, clavulanate allowed the classical β -lactam drug amoxicillin to be augmented in its antibacterial range. The combination of clavulanate and amoxicillin is called Augmentin (Fig. 4a) and has become front line therapy³². In parallel, the study of a penicillin analogue with a five-ring sulphur atom oxidized to the sulphone revealed that this derivative, called sulbactam, now had a weaker C–S bond that disposed the ring of the acyl-lactamase intermediate to open and create a long-lived covalent enzyme intermediate that was inactive³¹. The combination of sulbactam and ampicillin is Unasyn, which is also an important antibacterial combination³². Two other combinations of lactamase inactivator and β -lactam antibiotic, Timentin and Zocin, are also widely used antibacterial combinations that exemplify the utility of this approach.

Analogous logic could be used to screen for or design analogues of tetracyclines^{33,34} and erythromycins that would be less susceptible to recognition and transit by the efflux pumps. Third-generation forms of erythromycins, for example, the 16 ketolides³⁵ (Fig. 4b) that are currently in clinical testing, are less prone to induce the Erm type of methylation resistance. A combination strategy akin to the logic used for Augmentin is under study, where a separate inhibitor of efflux pumps, once identified as having sufficient potency and safety, could be combined with the macrolide antibiotic or the tetracycline.

When drug targets have been altered as in VRE, screening for molecules that are now active against the vancomycin-resistant strains have identified compounds such as LY333328³⁶ (Fig. 4c), a semisynthetic analogue of vancomycin. LY333328 contains a hydrophobic biphenyl substituent on the vancosamine sugar and is more hydrophobic and may partition the analogue more to the membrane³⁷, as well as alter its ratio of inhibition between transpeptidases and transglycosylases^{38,39}. Pristinamycin (trade name Synercid) is a combination of two non-ribosomal peptide natural products, quinupristin and dalbapristin, which act synergistically to inhibit protein synthesis in a bacteriocidal manner⁴⁰. The drug has recently been approved with the intent of treating VRE infections⁴¹.

Development of new classes of antibiotics

In addition to defining new targets in bacteria for new antibiotics (see below), there has been progress in development of a new structural class of synthetic molecules with broad spectrum and acceptable potency: the oxazolidinones. The oxazolidinones also inhibit protein biosynthesis, specifically by interaction with the 23S ribosomal RNA at or near the peptidyl transferase centre of the ribosome¹⁰. One such oxazolidinone, linezolid, has now progressed through clinical trials⁴² to approval in the United States. Linezolid has been described as the

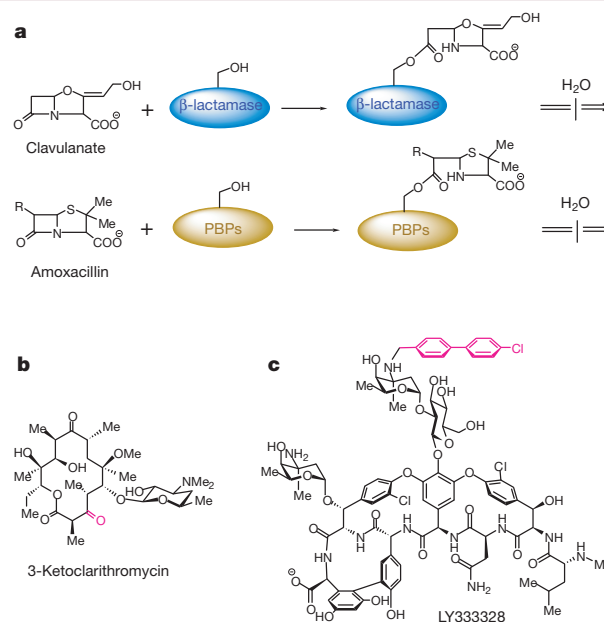


Figure 4 Counter-survival strategies. To combat the development of antibiotic resistance requires a knowledge of the specific mechanism of resistance and then development of a new class of compounds that are active against the resistant bacteria. **a**, Augmentin is a combination of clavulanate, which inactivates the β -lactamase by forming a slowly hydrolysing acyl enzyme intermediate, and amoxicillin, which blocks the cell wall-crosslinking transpeptidase, also by forming a slowly hydrolysing covalent acyl enzyme intermediate. **b**, Introduction of a 3-keto group into the macrolide ring in 3-keto clarithromycin (a ketolide) alters the conformation of the macrolactone and alters susceptibility to efflux and to induction of methylation resistance machinery. **c**, Screening against vancomycin-resistant enterococci (VRE) with semisynthetic variants of vancomycin indicates that hydrophobic substituents on the vancosamine sugar, as in LY333328 which has a biphenyl alkyl substituent on the amino sugar, restore about two logs of potency against VRE.

first new structural class of antibiotics introduced in three decades and will join the quinolones as evidence that synthetic compounds can be elaborated that match the antibiotic potency and selectivity of natural products. Also in development is a glycolipopeptide (17 amino acids cyclized to a macrolactone) called ramoplanin⁴³. This works against VRE by forming a complex with the lipid pentapeptide intermediates in cell-wall biosynthesis, acting in a somewhat analogous way to vancomycin by targeting a substrate rather than an enzyme in the peptidoglycan assembly pathway.

New approaches to antibiotic development and resistance

In this last section I take up the two questions: how do we find or make new antibiotics and can we slow down the development of resistance? The first question encompasses the searches both for new bacterial targets that will be killing sites and for new molecules with the killing (antibacterial) properties. The second question includes prescriptions for rethinking the ways that antibiotics are prescribed and used.

Using genomics to find new bacterial targets

Bacterial genomics has progressed to the extent that about three dozen bacterial genomes have been completely sequenced in the past few years, including the full gene sequence in such pathogens as *S. aureus*, *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, *Helicobacter pylori*, *Pseudomonas aeruginosa* and *Vibrio cholerae*. The genome of a prototypic streptomycete, *Streptomyces coelicolor*, which produces several types of polyketide antibiotics, is also close to being finished and should provide additional insights into the organization of antibiotic biosynthetic operons and their regulation. Approaches

involving gene disruption have begun to narrow the list of bacterial genes in pathogens that are essential either for virulence or for survival to perhaps a few hundred genes, some of known or putative function and some unknown. The products of these essential genes constitute an initial set of validated protein targets that can be loaded into optimized, automated screens. Chemical libraries can be assessed for inhibitors that score as hits, and these can then be elaborated into leads of useful potency and specificity, first *in vitro* and then in animal models of infections as antibacterial agents. These kinds of approaches offer significant promise for expansion of the current small number of targets that are the sites of action of contemporary antibiotics⁴⁴.

For example the enzyme that removes the formyl group from the amino-terminal formylmethionyl residue in bacterial proteins, peptide deformylase⁴⁵, is a metalloproteinase that is an essential gene product in many bacteria and for which potent inhibitors have recently been described⁴⁶. Several gene products that are manifestations of virulence responses by pathogenic bacteria⁴⁷ are being examined, including peptides involved in bacterial secretion pathways⁴⁸ and the two main types of signalling networks: two-component systems (transmembrane sensor kinase and response-regulatory gene-transcription factor)⁴⁹, and quorum sensing pathways that regulate selective gene activation in cell density-dependent responses^{50,51}.

The prospect of a much larger set of validated targets in bacteria will create opportunities for the development of novel therapeutics, such that the discovery of new molecules with antibiotic potential will gain in importance. Library approaches are being explored both with synthetic compounds and natural products. The synthetic libraries can be large, with millions of members, or focused around particular pharmacophores⁵² and contain a few hundred to thousand members in any iteration. Several of the chemical libraries seek to build in the identity and density of the functional groups and the three-dimensional architecture of compounds, demonstrating convergence towards the complexity of natural products⁵³.

The complementary approach of *in situ* library generation — combinatorial biosynthesis of natural products that act as antibiotics — is underway, especially for polyketide and mixed polyketide/non-ribosomal peptide antibiotics^{54–57} where all the biosynthetic genes are typically clustered together within a 50–100-kilobase region of DNA. This makes cloning of new biosynthetic operons for polyketides relatively straightforward, even from the large number of microbes that cannot at present be cultured in the laboratory environment. Meanwhile the multimodular organization of the lines of assembly of polyketide synthase enable domain and module swap, replacement and reprogramming strategies to make small libraries of modified polyketides by fermentation. Several erythromycin derivatives have been reported with up to three segments replaced combinatorially, which illustrates the potential of this approach⁵⁸. The rules need to be determined as to what kinds of point mutations, catalytic-domain replacements and multidomain module swaps can be tolerated and combined into a hybrid assembly line with good throughput of new products. For many of the macrolide antibiotics, represented by erythromycin, the sugars that are added to the macrocyclic lactone scaffolding by committed, tailoring enzymes are crucial determinants of biological activity. Efforts to provide alternate deoxy and amino sugars for *in vivo* alternate tailoring⁵⁹ have been reported and would be a second, multiplicative step for expanding the biosynthetic libraries of this class of antibiotics.

Strategies for extending antibiotic lifespan

Given that the emergence of antibiotic-resistant bacterial strains is an inevitable response to the widespread application of both current and future antibiotics, explicit rethinking of strategies for preserving and extending the useful life of antibacterial drugs has been central in this era of multiple drug-resistant bacterial recrudescence. Guidelines for altering the behaviour of both patients and physicians in antibiotic consumption and prescription have been advanced by

Table 1 Guidelines for extending the useful lifetimes of antibacterial drugs

- Optimal use of all antimicrobials
- Selective removal, control or restriction of antimicrobial agents or classes
- Use of antimicrobials in rotation or cyclic patterns
- Use of combination antimicrobial therapy to prevent the emergence of resistance

Adapted from the recommendations in ref. 60.

infectious disease specialists⁶⁰ (Table 1). In developed countries, the goal is to reduce the inappropriate prescription of antibiotics (estimates of up to 50% of prescriptions may be written for patients with viral infections that cannot respond to the antibacterial drugs). In developing countries, availability of antibiotics may be intermittent, the quality and potency uncertain, and the tendency to self-medicate without specialist input from physicians can be widespread. The net result is that individuals often take subtherapeutic doses of antibiotics; symptoms may disappear but resistant strains of bacteria are thereby selected and may only become acutely problematic to the individual in a subsequent medical crisis where an infection proves resistant. The threat to the larger population is that reservoirs of drug-resistant bacteria abound.

Among the guidelines considered to alter social norms about antibacterial drugs (Table 1) is that of rotating the use of antibiotics to preserve the efficacy of antibiotics of last resort, such as vancomycin in the treatment of life-threatening infections due to MRSA. The spectre of high-level vancomycin resistance transposed into MRSA is one of the doomsday scenarios in infectious disease⁶¹. There has also been the call to consider wider use of combinations of antibiotics in primary therapy, not just the Augmentin and Synercid approaches where two components work together to neutralize a single target, but combinations of distinct antibiotic classes that work on different targets concurrently⁶⁰. This combination approach has been the norm for curative regimens of anticancer drugs and more recently the multiple-component HAART (highly active antiretroviral therapy) regimens for control of AIDS progression⁶². But set against with the potential benefits of such an approach is the liability that if subtherapeutic doses are effected in combination therapy, multiple drug resistance will emerge more rapidly. The epidemic of multiple drug-resistant tuberculosis most probably arises from patient non-compliance during the extended period of drug therapy.

The use of antibacterial drugs as growth promoters for animals has been a topic of heated debate for well over a decade⁶³, as antibiotic prophylaxis in cattle, pigs, chicken and aquaculture grew to quantities that could dwarf human consumption of the same or related antibacterial drugs. As in humans, subtherapeutic doses in animals can select for resistant strains; if the bacteria cross from animal hosts to human hosts then reservoirs of resistance may markedly reduce the effective lifetime of human antibiotics. In Denmark in 1994, 24,000 kg of a vancomycin derivative called avoparcin was used for animal health, 1,000 times higher than the 24 kg of vancomycin used to treat human infections in the same year⁶³. When pigs treated with avoparcin were analysed for VRE strains, they contained the same five gene operons that encode vancomycin resistance in enterococci isolated from nonresponsive human patients. The Danish government later banned the use of avoparcin as an additive to animal feed. In a second example, VRE that were resistant to the recently approved quinupristin/dalfopristin combination were found to have arisen during drug treatment⁶⁴; this may have occurred through the effect of an animal reservoir involving an acetyl transferase carried by bacteria resistant to the related antibiotic virginiamycin, which has been used in animal feeds in Europe for 20 years⁶³. At the very least, principal drug classes, especially new ones that have therapeutic potential as human antibacterials should be carefully evaluated before being 'wasted' in animal feed uses.

In conclusion, the increased molecular knowledge about essential bacterial genes and the ability to screen such candidate targets with libraries of new synthetic and natural products to find hits that can be

subjected to iterative cycles of improvement of structure and function indicate that new antibacterial agents against non-traditional bacterial targets may be forthcoming. But new antibiotics by themselves will not alter the kinetics of the cycles of resistance development. Indeed, wider and more indiscriminate use could actually shorten the cycle time unless behaviour changes, which are difficult but not impossible to achieve, occur with regard to valuing antibiotics as precious and finite resources. □

1. Neu, H. C. The crisis in antibiotic resistance. *Science* **257**, 1064–1073 (1992).
2. Gold, H. S. & Moellering, R. C. Antimicrobial-drug resistance. *N. Engl. J. Med.* **335**, 144–1453 (1996).
3. Levy, S. B. The challenge of antibiotic resistance. *Sci. Am.* **275**, 46–53 (1998).
4. Levy, S. B. Multidrug resistance—a sign of the times. *N. Engl. J. Med.* **338**, 1376–1378 (1998).
5. Spratt, B. G. & Cromie, K. D. *Rev. Infect. Dis.* **10**, 699 (1988).
6. Williams, D. H. The glycopeptide story—how to kill the deadly “superbugs”. *Nat. Prod. Rep.* **13**, 469–477 (1996).
7. Brisson-Noel, A., Trieu-Cuot, P. & Courvalin, P. Mechanism of action of spiramycin and other macrolides. *J. Antimicrob. Chemother.* **22**(Suppl. B), 13–23 (1988).
8. Chopra, I. in *The Tetracyclines, Handbook of Experimental Pharmacology* Vol 78 (eds Hlavaka, J. J. & Boothe, J. H.) 317–392 (Springer, Berlin, 1985).
9. Fourmy, D., Recht, M. L., Blanchard, S. C. & Puglisi, J. D. Structure of the A site of the *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274**, 1371–1376 (1996).
10. Kloss, P., Xiong, L., Shinabarger, D. L. & Mankin, A. S. Resistance of mutations in 23S rRNA identify the site of action of the protein synthesis inhibitor linezolid in the ribosomal peptidyl transferase center. *J. Mol. Biol.* **294**, 93–101 (1999).
11. Shen, L. L. in *Quinolone Antibacterial Agents*, 2nd edn (eds Hooper, D. C. & Wolfson, J. S.) 77 (American Society for Microbiology, Washington DC, 1993).
12. Maxwell, A. DNA gyrase as a drug target. *Trends Microbiol.* **5**, 102–109 (1997).
13. Ferrero, L., Cameron, B. & Crouzet, J. Analysis of *gyrA* and *griA* mutations in stepwise-selected ciprofloxacin-resistant mutant of *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **39**, 1554–1558 (1995).
14. Davies, J. Bacteria on the rampage. *Nature* **383**, 219–220 (1996).
15. Murray, E. Vancomycin resistant enterococci. *Am. J. Med.* **102**, 284–293 (1997).
16. Arthur, M. & Courvalin, P. Genetics and mechanisms of glycopeptide resistance in enterococci. *Antimicrob. Agents Chemother.* **37**, 1563–1571 (1993).
17. Walsh, C., Fisher, S. L., Park, I. S., Prahalad, M. & Wu, Z. Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem. Biol.* **3**, 21–28 (1996).
18. Davies, J. Inactivation of antibiotics and the dissemination of resistance genes. *Science* **264**, 375–382 (1994).
19. Schentag, J. J. *et al.* Genesis of methicillin resistant *Staphylococcus aureus* (MRSA). *Clin. Infect. Dis.* **26**, 1204–1212 (1998).
20. Jacoby, G. A. & Archer, G. L. New mechanisms of resistance to antibacterial agents. *N. Engl. J. Med.* **324**, 601–612 (1991).
21. Levy, S. B. Active efflux mechanisms for antimicrobial resistance. *Antimicrob. Agents Chemother.* **36**, 695–703 (1992).
22. Paulsen, I. T., Brown, M. H. & Skurray, R. A. Proton-dependent multidrug efflux systems. *Microbiol. Rev.* **60**, 575–608 (1996).
23. Ross, J. *et al.* Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super gene family. *Mol. Microbiol.* **4**, 1207–1214 (1990).
24. Philippon, A., Labia, R. & Jacoby, G. Extended spectrum beta-lactamases. *Antimicrob. Agents Chemother.* **28**, 302–307 (1985).
25. Shaw, K. J., Rather, P. N., Hare, S. R. & Miller, G. H. Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol. Rev.* **57**, 138–163 (1993).
26. Hon, W. C. *et al.* Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. *Cell* **89**, 887–895 (1997).
27. Bussièr, D. *et al.* Crystal structure of ErmC, an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* **37**, 7103–7112 (1998).
28. Bugg, T. D. H. *et al.* Molecular basis for vancomycin resistance in *Enterococcus faecium*, BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. *Biochemistry* **30**, 10408–10415 (1991).
29. Song, M. D., Wachi, M., Doi, M., Ishino, F. & Matsushashi, M. Evolution of an inducible penicillin target protein in MRSA by gene fusion. *FEBS Lett.* **221**, 167 (1987).
30. Spratt, B. G. Resistance to antibiotics mediated by target alterations. *Science* **264**, 388–393 (1994).
31. Knowles, J. R. Penicillin resistance: the chemistry of beta-lactamase inhibition. *Acc. Chem. Res.* **18**, 97–105 (1985).
32. The choice of antibacterial drugs. *Med. Lett.* **41**, 95–104 (1999).
33. Nelson, M. L., Park, B. H. & Levy, S. B. Molecular requirement for the inhibition of the tetracycline antiport protein and the effect of potent inhibitors on the growth of tetracycline-resistant bacteria. *J. Med. Chem.* **37**, 1355–1361 (1994).
34. Chu, D. W., Plattner, J. J. & Katz, L. New directions in antibacterial research. *J. Med. Chem.* **39**, 3853–3874 (1996).
35. Agouridas, C., Benedetti, Y., Denis, A., Le Martert, O. & Chantot, J. F. Ketolides: a new distinct class of macrolide antibacterials, synthesis and structural characteristics of RU 004. *35th Interscience Conference on Antimicrobial Agents and Chemotherapy* Abstr. F158 (American Society for Microbiology, 1995).
36. Biavasco, F. *et al.* In vitro antibacterial activity of LY333328, a new semisynthetic glycopeptide. *Antimicrob. Agents Chemother.* **41**, 2165–2172 (1997).
37. Williams, D. H., Maguire, A. J., Tsuzuki, W. & Westwell, M. S. An analysis of the origins of cooperative binding energy of dimerization. *Science* **280**, 711–713 (1998).
38. Ge, M. *et al.* Vancomycin derivatives that inhibit peptidoglycan biosynthesis without binding D-Ala-D-Ala. *Science* **284**, 507–511 (1999).
39. Walsh, C. Deconstructing vancomycin. *Science* **284**, 442–443 (1999).
40. Barry, A. L. & Fuchs, P. C. In vitro activities of a streptogramin (RP59500), three macrolides, and an azolid against four respiratory tract pathogens. *Agents Chemother.* **39**, 238–240 (1995).
41. Linden, P. K., Pasculle, A. W., McDevitt, D. & Kramer, D. J. Effect of quinupristin/dalfopristin (synercid) treatment of infections caused by vancomycin-resistant *Enterococcus faecium* (VREF). *J. Antimicrob. Chemother.* **39**(Suppl. A), 145–151 (1997).
42. Chien, J. W., Kucia, M. L. & Salata, R. A. Use of linezolid, an oxazolidinone, in the treatment of multidrug-resistant gram-positive bacterial infections. *Clin. Infect. Dis.* **30**, 146–151 (2000).
43. Kurz, M. & Guba, W. 3D structure of ramoplanin: a potent inhibitor of bacterial cell wall biosynthesis. *Biochemistry* **35**, 12570–12575 (1996).
44. Chapra, I., Hodgson, J., Metcalf, B. & Poste, G. The search for antimicrobial agents effective against bacteria resistant to multiple antibiotics. *Antimicrob. Agents Chemother.* **41**, 497–503 (1997).
45. Chan, M. K. *et al.* Crystal structure of *E. coli* peptide deformylase. *Biochemistry* **36**, 13904–13909 (1997).
46. Chen, D. Z. *et al.* Actinonin, a naturally occurring antibacterial agent is a potent peptide deformylase inhibitor. *Biochemistry* **39**, 1256–1262 (2000).
47. Barret, J. F. & Isaacson, R. E. Bacterial virulence as a potential target for therapeutic intervention. *Annu. Rev. Med. Chem.* **30**, 111–118 (1995).
48. Stephens, C. & Shapiro, L. Bacterial protein secretion—a target for new antibiotics? *Chem. Biol.* **4**, 637–641 (1997).
49. Barret, J. F. & Hoch, J. A. Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob. Agents Chemother.* **42**, 1529–1536 (1998).
50. Kleerebezem, M., Quadri, L. E. N., Kuipers, O. P. & de Vos, W. M. Quorum sensing by peptide pheromones and two-component signal-transduction systems in gram-positive bacteria. *Mol. Microbiol.* **24**, 895–904 (1997).
51. Reimann, C. *et al.* The global activator *gacA* of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Mol. Microbiol.* **24**, 309–319 (1997).
52. Sofia, M. J. *et al.* Discovery of novel disaccharide antibacterial agents using a combinatorial library approach. *J. Med. Chem.* **42**, 8194–8198 (1999).
53. Tan, D. S., Foley, M. A., Shair, M. D. & Schreiber, S. L. Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays. *J. Am. Chem. Soc.* **120**, 8565–8566 (1998).
54. Leadlay, P. F. Combinatorial approaches to polyketide biosynthesis. *Curr. Opin. Chem. Biol.* **1**, 162–168 (1997).
55. Katz, L. Manipulation of modular polyketide synthases. *Chem. Rev.* **97**, 2557–2575 (1997).
56. Khosla, C. Harnessing the biosynthetic potential of modular polyketide synthases. *Chem. Rev.* **97**, 2577–2590 (1997).
57. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68 (1998).
58. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel “unnatural” natural products. *Proc. Natl Acad. Sci. USA* **96**, 1846–1851 (1999).
59. Zhao, L. *et al.* Engineering a methymycin/pikromycin-calicheamicin hybrid: construction of two new macrolides carrying a designed sugar moiety. *J. Am. Chem. Soc.* **121**, 9881–9882 (1999).
60. Schlaes, D. M. *et al.* Guidelines for the prevention of antimicrobial resistance in hospitals. *Clin. Infect. Dis.* **25**, 583–599 (1997).
61. Waldvogel, F. A. New resistance in *Staphylococcus aureus*. *N. Engl. J. Med.* **340**, 556–557 (1999).
62. Drugs for HIV infection. *Med. Lett.* **42**, 1–6 (2000).
63. Witte, W. Medical consequences of antibiotic use in agriculture. *Science* **279**, 996–997 (1998).
64. Chow, J. W., Donahedian, S. M. & Zervos, M. J. Emergence of increased resistance to quinupristin/dalfopristin during therapy for *Enterococcus faecium* bacteremia. *Clin. Infect. Dis.* **24**, 90–91 (1997).

Acknowledgements

Experimental work cited in this review is supported by the National Institutes of Health. I thank R. Chen and I. Lessard for preparation of the artwork.

Toll-like receptors in the induction of the innate immune response

Alan Aderem* & Richard J. Ulevitch†

*Department of Immunology, University of Washington, Seattle, Washington 98195, USA

†Department of Immunology, The Scripps Research Institute, La Jolla, California 92037, USA

The innate immune response is the first line of defence against infectious disease. The principal challenge for the host is to detect the pathogen and mount a rapid defensive response. A group of proteins that comprise the Toll or Toll-like family of receptors perform this role in vertebrate and invertebrate organisms. This reflects a remarkable conservation of function and it is therefore not surprising that studies of the mechanism by which they act has revealed new and important insights into host defence.

The immune response to microbial pathogens relies on both innate and adaptive components¹. The immediate, innate response is mediated largely by white blood cells such as neutrophils and macrophages, cells that phagocytose and kill the pathogens, and that concurrently coordinate additional host responses by synthesizing a wide range of inflammatory mediators and cytokines². In macrophages, the infectious agent is killed and degraded within the maturing phagosome, and components of the pathogen are presented to T cells, resulting in the activation of the adaptive immune response and the establishment of protective immunity². A primary challenge to the innate immune system is the discrimination of a large number of potential pathogens from self, with the use of a restricted number of receptors. This problem is compounded by the tendency of pathogens to mutate. This challenge has been met by the evolution of a variety of receptors that recognize conserved motifs on pathogens that are not found in higher eukaryotes. These motifs have essential roles in the biology of the invading agents, and are therefore not subject to high mutation rates. Janeway & Medzhitov³ have provided a set of definitions to formalize a description of the components of the innate immune system. They propose calling the motifs pathogen-associated molecular patterns (PAMPs), and their cognate binding partners on the phagocytes pattern-recognition receptors³. Pathogen-associated motifs include mannans in the yeast cell wall, formylated peptides and various bacterial cell-wall components such as lipopolysaccharide (LPS), lipopeptides, peptidoglycans and teichoic acids³ (Table 1). There are two principal classes of pattern recognition receptor: those that mediate phagocytosis and those that lead to the activation of pro-inflammatory pathways^{2,3}.

The best studied model of innate immunity is one using Gram-negative bacteria and bacterial endotoxin (LPS)⁴. LPS is considered to play a key role in the septic shock syndrome in humans owing to its profound effects on the innate immune system. It is therefore not surprising that studies of the pathogenesis of septic shock has time and again provided the key insights into the molecular basis of innate immunity. We discuss many of these recent advances below.

LPS is the prototypic activator of innate immunity; it is active at concentrations below 1 nM, and studies with LPS have revealed the sequence and mechanisms of the essential

responses⁴. Despite this focus on LPS, it is now clear that other microbial products, structurally distinct from LPS, act as potent activators of the innate immune system. It is therefore useful to consider briefly the various classes of these activating substances, and then discuss them in view of what is known at the mechanistic level. But although it has been useful to analyse the pathways leading to immune activation triggered by purified microbial components, it is important to remember that the immune system encounters a 'cocktail' of these molecules when interacting with pathogens *in vivo*.

Components of pathogens that activate immunity

Within the known phyla of bacteria, there are four groups of especially significant human pathogens. These are Gram-negative and Gram-positive bacteria, mycobacteria and spirochaetes. A wide variety of bacterial components are capable of stimulating innate immune responses. These include LPS, peptidoglycan, lipoteichoic acid, lipoarabinomannan (LAM), lipopeptides and bacterial DNA (see Table 1). LPS is the principal component of Gram-negative bacteria that activates the innate immune system. The chemical structure of these complex outer-membrane glycolipids has been studied extensively and discussed in numerous reviews^{5,6}. The essential structural feature of LPS that governs interactions with the innate immune system is known as lipid A (ref. 6). Regardless of the type of Gram-negative bacterium, the lipid A is composed of a diglucosamine backbone containing ester-linked and amide-linked long-chain fatty acids. Recent results suggest that the fatty-acid composition might dictate the binding specificities of key receptors^{7,8}. Bacterial lipoproteins are also potent stimulators of inflammatory responses, and are found in all Eubacteria⁹. For example, lipoproteins from mycobacteria, mycoplasma and spirochaetes activate macrophages. These outer-membrane lipoproteins contain a lipid-modified cysteine residue at the amino terminus; this lipid-containing domain is essential for the biological effects of the lipoproteins⁹. Gram-positive bacteria contain cell walls composed of a number of potentially biologically active molecules, including peptidoglycan, lipoteichoic acid and lipopeptides⁹. LAM is a key stimulator of innate immunity found in mycobacteria; it and a variety of peptidoglycan and lipopeptide species are the molecules that endow complete Freund's adjuvant with its efficacy¹⁰. Among the spirochaetes only *Leptospira* have LPS in the outer membrane¹¹. Although leptospiral LPS might be an

important component in this group of bacteria, current opinion supports the contention that outer-membrane lipoproteins are the principal activating signal to the innate immune system during infection by spirochaetes¹².

For other microbial pathogens such as yeast or fungi and protozoan parasites, the spectrum of molecules known to stimulate innate immune responses are further extended. For example, in *Candida albicans* cell-wall constituents such as glucan, chitin, mannans and mannoproteins are potent immune modulators¹³, whereas in *Plasmodium falciparum* glycosylphosphatidylinositol molecules are stimulatory¹⁴. Interestingly, not all cell-wall constituents activate innate immune responses; lipophosphoglycan in *Leishmania donovani* inhibits proinflammatory responses in macrophages¹⁵. It will be important to continue to characterize the structure of additional molecules present in the outer membranes and/or cell walls of pathogens that activate the innate immune system. As discussed below, there are at least nine members of the Toll family and only for two of them have ligands been defined, suggesting that many ligand–receptor pairs remain to be identified.

Recognition of molecular motifs on pathogens

Studies of the pathogenesis of Gram-negative septic shock led to a new level of understanding of mechanisms of innate immunity¹⁶. First came the discovery and characterization of LPS-binding protein (LBP); subsequently the identification of the importance of CD14 resulted in studies that formed the framework for our current state of knowledge^{17,18}. Extensive biochemical, genetic and animal-model data support the concept that activation of host defence mechanisms when Gram-negative bacteria are present requires the recognition of LPS by mechanisms dependent on LBP and CD14 (ref. 16) (Fig. 1). Thus, LBP is an opsonin and CD14 is an opsonic receptor for complexes of LPS (or LPS-containing particles such as bacteria) and LBP. CD14 has been implicated in cell activation processes involving other products of microbial pathogens including LAMs, peptidoglycans

and outer-membrane lipoproteins¹⁶. The question of the involvement of opsonic proteins for these varied stimuli has not been answered and remains to be addressed in detail. Moreover, there are few, if any, relevant details about the structure of CD14 that provide insight into the ligand-binding pocket of this protein. The fact that both CD14 and members of the Toll family contain multiple leucine-rich repeats suggests novel ligand-binding sites as well as protein–protein interactions at the cell surface that are not yet understood. Numerous studies support the contention that CD14 functions solely as a ligand-binding protein (LPS or LPS–LBP complexes) that does not participate in the generation of a transmembrane signal. Rather, many workers hypothesized that one or more additional transmembrane proteins acted in concert with LPS–CD14 complexes to initiate the signalling processes leading to cell activation. Extensive efforts to identify this putative transmembrane protein were unsuccessful until very recently.

Approximately 8 years after the initial publications delineating the importance of LBP and CD14 came the next main advance in understanding innate immunity: the identification of the putative transmembrane protein that acted with CD14 to generate a transmembrane signal linked to LPS-induced cell activation. The impetus for this advance was two seminal discoveries: the role of the Toll-like receptors (TLRs) in innate immunity in *Drosophila*^{19,20}, and the identification of a TLR homologue as the gene responsible for LPS responses in two natural mouse mutants^{21–23}. These results marked the beginning of what will most certainly be a large leap forward in understanding how innate immune responses function in regulating responses to infection with microbial pathogens. They are of equal importance to the discoveries underlying our understanding of plasma-membrane receptors that control adaptive immune responses.

The TLR family

Although *Drosophila* has no adaptive immune system, it is very resistant to microbial infections²⁴. This is because the innate immune

Table 1 PAMPs and pattern recognition receptors

PAMP	Pathogen(s)	Essential ligand	Pattern recognition receptors	Biological sequelae
LPS	Most Gram-negative bacteria	Lipid A	LBP, CD14 TLR4, TLR2* Scavenger receptor	Enhance inflammatory response mediated by TLRs Recognize LPS and initiate inflammatory response Endocytosis of LPS (non-inflammatory), phagocytosis
Lipoproteins	Eubacteria	Amino-terminal tripalmitoylated cysteine† generated by the signal peptidase II cleavage system	TLR2	Initiates inflammatory response
Peptidoglycan	Most bacteria	Undefined	CD14 TLR2	Enhances inflammatory response Initiates inflammatory response
Lipoteichoic acid	Many Gram-positive bacteria	Undefined	TLR2, TLR4‡	Initiates inflammatory response
CpG	Many microbial pathogens	Unmethylated CpG-containing oligonucleotide§	Undefined	Initiates inflammatory response
Lipoarabinomannan	Mycobacteria	Undefined	TLR2 CD1	Initiates inflammatory response Presents glycolipid to αβ T cells
N-formyl-Met	Prokaryotes	Amino-terminal N-formylmethionine of proteins synthesized <i>de novo</i>	f-Met receptors 1 and 2	Chemotaxis and release of inflammatory mediators (?)
Mannans and mannoproteins	Yeast	Undefined	Mannose receptor Mannose-binding protein	Phagocytosis, endocytosis and initiation of inflammatory response Opsonization and complement fixation
Zymosan (yeast cell wall)	Yeast	Undefined	Mannose and β-glucan receptors TLR2	Phagocytosis Initiates inflammatory response
Heat shock proteins	Prokaryotes and eukaryotes	Undefined	Undefined	Initiate inflammatory response and promote T cell dependent immune responses

* TLR2's role as an LPS receptor is based on studies *in vitro*, and, although TLR2 may function as a physiological receptor for LPS, evidence to support this function *in vivo* has not yet been demonstrated.

† Mycoplasma lipoproteins have been shown to have a dipalmitoylated amino-terminal cysteine residue, which is also recognized by TLR2. These findings indicate that not all three fatty acid ester-linked chains are necessary for recognition.

‡ LTA is recognized by TLR2 in systems *in vitro*, and by TLR4 in studies with knockout mice. These differences have not been reconciled.

§ Flanking sequences are also influential.

|| Studies with Hsp60 family members have shown that C3H/HeJ mice are defective in inflammatory signalling in response to these proteins, implicating TLR4 as a pattern recognition receptor for foreign and endogenous Hsp60 family gene products.

system provides protection that is in part due to the synthesis of potent antimicrobial peptides. These peptides are induced in response to signalling pathways activated by at least two members of the TLR family found in *Drosophila*, dToll and 18-wheeler. The activation of dToll induces an anti-fungal peptide, drosomycin, whereas the ligation of 18-wheeler induces an antibacterial peptide, attacin^{19,20,25}. An essential step in the activation of such responses in *Drosophila* seems to be the activation of a proteolytic cascade that produces peptidic ligands for the Toll family of receptors. Whether or not this mechanism is unique to *Drosophila* or has been conserved in mammalian cells remains unanswered. Remarkably, these studies suggest that the TLRs were capable of discriminating between fungi and bacteria, and further that they were capable of inducing an appropriate and distinct antimicrobial response. Activation of these pathways in *Drosophila* via the membrane receptor initiates an intracellular kinase cascade that ultimately produces a translocation of transcription factors, Dif and Relish²⁴, from cytoplasm to nucleus. Dif and Relish are homologous to NF- κ B, a transcription factor known to activate a variety of inflammatory mediators and cytokines including tumour necrosis factor- α (TNF- α) and interleukin-12 (IL-12)²⁵, but it is still not clear what differences in signalling downstream of dToll and 18-wheeler account for the divergent patterns of gene expression. Because it is likely that mammalian cell activation via different TLRs produces distinct patterns of gene expression, the continued study of innate immunity in *Drosophila* is likely to provide new clues about related processes in mammalian cells.

Janeway and colleagues, in the search for receptors within the innate immune system, began a search for dToll-related proteins. Their efforts provided the next key advance in our understanding of innate immunity: they identified the first human homologue of *Drosophila* toll, initially termed human Toll and subsequently termed TLR4 (ref. 26). This observation suggested a probable link to innate immunity in general because a constitutively active mutant of TLR4 activated NF- κ B-controlled genes such as IL-1, IL-6 and IL-8 (ref. 26). Importantly, constitutively active TLR4 also induced members of the B7 family, molecules that are required for the activation of naive T cells by antigen-presenting cells²⁶. TLR4 was therefore a potentially important link between pathogen detection and the induction of the adaptive immune response. It was reasonable to wonder whether other members of this family were present in mammalian cells; soon afterwards, four more human TLRs (hTLRs) were discovered. Subsequent studies by others provided a detailed analysis of the structural features that link these proteins to *Drosophila* toll and to the IL-1 receptor family²⁷. Importantly, Rock *et al.* highlighted the general structural features of the TLR family, namely the presence of multiple leucine-rich repeats in the ectodomain and the Toll-homology domain found in the cytoplasmic tail of all members of this protein family.

Establishing a link between TLRs and pathogen recognition by means of CD14-dependent mechanisms was essential to advance our understanding of the possible role of TLRs in innate immunity. Progress came rapidly as several groups established model systems for defining the function of one TLR, TLR2. First, it was demonstrated that the ectopic expression of TLR2 would support the LPS-induced activation of NF- κ B in the transfected cells^{28,29}. Second, these model systems permitted an analysis of downstream signalling pathways. Importantly, both groups provided data to support the contention that TLR2 signalling to NF- κ B involves steps that are similar to those used by the IL-1 receptor^{28,29}. These common steps include the involvement of the adaptor protein MyD88, the serine kinase IL-1R-associated kinase (IRAK) and another adaptor protein, TBF-receptor-associated factor 6 (TRAF6). It must be kept in mind that these initial reports did not document a physiological role for TLR2 as an LPS receptor. In fact the function of TLR2 as an LPS receptor remains unclear. As noted below, genetic data strongly support the contention that TLR4 is the predominant, if not the exclusive, receptor for LPS isolated from most Gram-negative organisms.

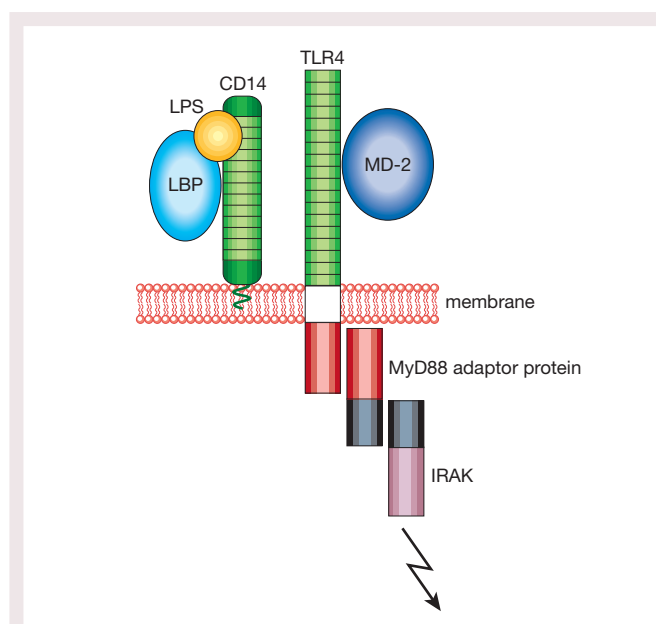
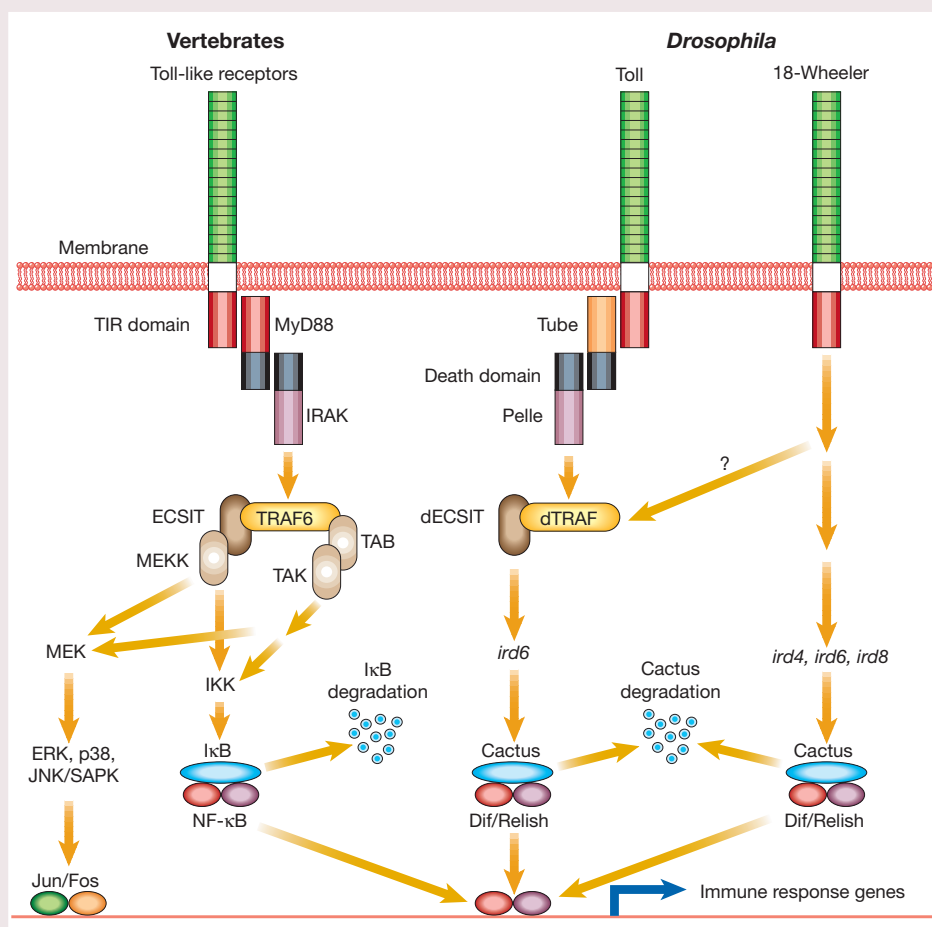


Figure 1 Recognition of LPS on the surface of phagocytes. LPS is opsonized by LBP, and the complex is recognized by the opsonic receptor, CD14, on the macrophage surface. CD14 associates with the cell surface by means of a glycolipid linkage and is not capable of generating a transmembrane signal. It is likely that the LPS-LBP-CD14 ternary complex activates TLR4 in some way, which in turn signals through the adaptor protein MyD88 and the serine kinase IRAK. The mechanism by which TLR4 is activated is not known and could be either direct or indirect. Both TLR4 and CD14 contain multiple leucine-rich repeats (brown); and these domains might facilitate protein-protein interaction. MD-2 is a secreted protein that binds to the extracellular domain of TLR4 and is important in its signalling.

Again, the many efforts to understand the pathogenesis of septic shock paved the way to our current understanding of how LPS is recognized in physiological settings. For more than two decades it had been understood that certain strains of mice were resistant to LPS and that this resistance was a consequence of a genetic defect. The seminal work of Beutler and co-workers demonstrating that the genetic defect in two strains of mice that are hypo-responsive or non-responsive to LPS was linked to TLR4 (refs 21, 22) led to a firm conclusion about how LPS is recognized. The co-dominant *Lpsd* allele of the C3H/HeJ strain was a result of a mis-sense mutation in the third exon of TLR4; this mutation was predicted to result in a Pro 712→His substitution^{21,22}. When this mutation was introduced into wild-type TLR4, the receptor was converted into a dominant-negative mutant that inhibited LPS-dependent responses in a transfected macrophage cell line³⁰. Another LPS-resistant strain, C57Bl/10ScCr, was shown to be homozygous for a null mutation of TLR4 (refs 21, 22). These findings provided the first direct link between the TLRs and physiological responses to LPS. This contention was supported by the demonstration that TLR4-null mice had a phenotype that was similar to that of the C3H/HeJ strain²³. Another important finding supporting an essential role for TLR4 in LPS-dependent responses was provided by the observation that Chinese hamsters respond normally to LPS even though they carry a null allele for TLR2 (ref. 31). Finally it was shown that a dominant-negative mutant of TLR2 did not effect LPS responsiveness in transfected macrophages³⁰. Thus a large gap in our knowledge about how LPS is recognized was closed by the combined efforts of many investigators with a primary interest in defining the mechanisms responsible for septic shock.

Receptors of the acquired immune system are composed of multi-protein complexes that permit the tight regulation of cellular activation through a variety of mechanisms. It is therefore not

Figure 2 Signalling pathways activated by TLRs in vertebrates and in *Drosophila*. The TLRs have an intracellular domain that is homologous with that of the IL-1 receptor, and is known as TIR. TIR binds to a homologous domain in an adaptor protein, MyD88, which also contains a death domain; this interacts with a death domain in the serine kinase IRAK. IRAK interacts with an adaptor known as TRAF6. TRAF6 links to the MAP 3-kinase TAK-1, through an adaptor TAB2. TAK-1 is involved in the activation of the transcription factor NF- κ B through the activation of I κ B kinases, and in the activation of the AP-1 transcription family members Jun and Fos, by way of additional MAP kinases. Both AP-1 transcription family members and NF- κ B are required for the transcription of immune response genes. TRAF6 is known to act through more than one pathway. For example, the adaptor ECSIT (evolutionarily conserved intermediate in Toll pathways) bridges TRAF6 to the MAP 3-kinase MEKK-1. The *Drosophila* Toll pathway is similar to that of the vertebrate TLRs. Toll links to an adaptor tube, the functional homologue of MyD88. Tube binds the kinase Pelle, a homologue of IRAK. A number of other homologues of the vertebrate pathways are found in *Drosophila*, including dTRAF and dECSIT, although it is not yet clear where precisely they fit into the signalling pathway. Similarly, additional *Drosophila* genes, including *ird4*, *ird6* and *ird8*, have been found through genetic screens, although precisely how they fit into the toll signalling pathway is not yet known. Finally, Cactus is a homologue of I κ B, whereas Dif and Relish are homologues of NF- κ B. See the text for a more complete description of the signalling pathway.



surprising that receptors that regulate innate immune responses might have similar degrees of complexity. This likelihood is best illustrated by the current level of understanding of the LPS receptor. The importance of CD14 and TLR4 is well established. Most recently an additional protein, termed MD-2, has been shown to be important for signalling via TLR4 (ref. 32). MD-2 is a secreted protein that apparently functions by binding to the extracellular domain of TLR4, where it facilitates LPS responsiveness, perhaps by stabilizing TLR4 dimers. However, the precise function of MD-2 in the LPS receptor complex is unknown at present. Defining the function of MD-2 in terms of its role in supporting LPS signalling as well as determining the interrelations of the proteins of the LPS receptor remain as important questions for future research.

The importance of TLR2 as an LPS receptor still remains unresolved, although the genetic data strongly support the contention that TLR4-dependent signalling mechanisms predominate. Is there a species difference? The results from studies *in vitro* implicating TLR2 in LPS signalling were obtained with the human receptor, whereas the genetic studies were performed in the mouse. This latter argument seems unlikely in view of a very recent report that an antibody against TLR2 does not inhibit LPS-induced TNF- α production in human monocytes, although it inhibits TLR2 activation by heat-killed *Listeria monocytogenes*³³. It seems likely that TLR2 can function as an LPS receptor when overexpressed but that it is not important for LPS responses *in vivo*. It remains to be determined whether or not TLR2 functions as a receptor for LPS from other organisms, such as spirochaetes or from other bacteria with lipid A of atypical structure. In contrast, a substantial number of reports detailing biochemical and genetic data provide strong support for the contention that

TLR2, but not TLR4, is a receptor for components of Gram-positive bacteria, mycobacteria, yeast and other microbial pathogens. Thus, TLR2-null mice and TLR2-null Chinese hamster ovary cells respond to LPS but not to Gram-positive bacterial component^{34,35}. Similarly, macrophages expressing a dominant-negative mutant of TLR2 do not respond to Gram-positive bacteria but are fully functional when challenged with LPS³⁰. More or less simultaneously, several groups provided evidence that Gram-positive cell-wall components, including peptidoglycan and lipoteichoic acid, activate cells via TLR2, and that TLR2 signals in response to lipopeptides from a wide variety of bacteria³⁴⁻³⁷. However, whereas TLR4 seems to be selective for LPS (although there is an example of its signalling the response to HSP60 (ref. 38)), TLR2 is much more promiscuous. Not only does TLR2 signal in response to different chemical structures such as peptidoglycan and lipopeptide³⁵⁻³⁷, but it is also required for pro-inflammatory signalling to mycobacterial cell-wall components including lipopeptides, lipoarabinomannan and mycolylarabinogalactan-peptidoglycan complexes³⁹⁻⁴². In addition, TLR2 is required for the pro-inflammatory pathways stimulated in macrophages by zymosan, a yeast cell-wall preparation³⁰. Taken together, these results support the contention that TLR4 functions as a receptor for LPS from Gram-negative bacteria and TLR2 is involved in the recognition of multiple products of Gram-positive organisms, mycobacteria and yeast. This ability of TLR4 and TLR2 to discriminate between pathogens is remarkably similar to the selectivity between pathogens observed for the *Drosophila* TLRs, dToll and 18-wheeler^{19,20} described above. One important question that has yet to be answered is the sequence of events that occur after TLR2 or TLR4 is engaged by a physiological ligand. This includes downstream signalling molecules and perhaps

most importantly a description of the genes that are induced as a result of activation of these receptors.

So far, nine different mammalian TLRs have been described^{27,43,44} and the complete sequence of TLR10 has been elucidated (T.-H. Chuang and R.J.U., unpublished observations). It is not unreasonable to assume that additional TLRs might be found as genomic databases become more complete. So far, genetic data suggest that the TLRs have unique functions and are not redundant. The identification of ligands for the TLRs is therefore an important area of research. It is safe to assume that many activators of the innate immune system have not yet been paired with TLR family members; we expect progress with this soon. For example, recent work has implicated TLRs in pro-inflammatory responses induced by bacterial DNA (A. Ozinsky *et al.*, unpublished observations). More interesting perhaps are the observations that different PAMPs are recognized by distinct combinations of TLRs, suggesting that TLRs can establish a combinatorial repertoire to discriminate between the large number of PAMPs found in nature (A. Ozinsky *et al.*, unpublished observations). In fact the number of possibilities for pairing of TLRs is enormous. Moreover, one must also consider the possibility of the existence of other MD-2-like molecules that might regulate specific functions of TLRs.

The phagocytosis of bacteria and other pathogens by macrophages is one of the initiating events of the innate immune response. During phagocytosis, TLRs are recruited to the phagosomes, where they sample the contents and determine the nature of the pathogen³⁰. Thus specific TLRs might distinguish between components in the phagosome and participate in the formulation of an inflammatory response appropriate for defence against a specific pathogen.

It is still unknown whether the TLRs directly bind PAMPs in mammalian cells. In *Drosophila* this seems to occur indirectly: microbial pathogens activate a protease cascade that generates a peptide ligand for the receptor¹. The ligand of dToll has been defined as a proteolytic cleavage product named spaetzle⁴⁵. It is conjectured that this polypeptide binds to dToll although there are at present no direct data to prove this. Moreover, the ligand for 18-wheeler has not yet been identified. Does such a model describe events at the cell surface of myeloid lineage cells that have been exposed to bacterial LPS? This seems unlikely but it must still be considered. Two recent reports suggest that LPS binds TLR4 directly^{7,8}. As mentioned above, the active moiety of LPS is a glycolipid known as lipid A. Lipid IVa is a partial structure of lipid A with the peculiar property that it acts as an agonist of pro-inflammatory responses in the mouse and as an antagonist in humans. Transfection of mouse (or hamster) TLR4 into human macrophages gives them the ability to detect lipid IVa as an agonist, whereas transfection of mouse macrophages with human TLR4 has the opposite result^{7,8}. Thus TLR4 alone determines how the macrophage responds to lipid IVa. This is hard to reconcile with the *Drosophila* model with an upstream proteolytic cascade. Perhaps the proteolytic cascade found in *Drosophila* is reflected by the complement and coagulation cascades of mammalian species. These latter systems are known to be activated directly by pathogens releasing peptidic products that act via receptor systems distinct from the TLRs.

TLR signalling pathways

Signalling via the TLRs has been studied by several groups, primarily in systems dependent on TLR2 and TLR4^{26,46–49}. A framework that has guided studies of TLR signalling was provided by findings from investigations of signalling via the related IL-1 receptor (IL-1R) family^{50,51}. The similarities between these two classes of receptor no doubt derive from the common Toll homology domain present in the cytoplasmic tails. Most studies have focused on the activation of TLRs leading to NF- κ B translocation and transactivation, although other events such as the activation of the mitogen-activated protein (MAP) 3-kinase pathway are also triggered after TLR ligation. Both genetic and biochemical data support the scheme shown in Fig. 2 and discussed below.

Ligation of a TLR promotes dimerization and results in the recruitment of MyD88, which contains two domains: a C-terminal Toll homology domain that interacts with the Toll homology domain of the receptor, and an N-terminal death domain^{26,48}. This death domain undergoes homophilic interaction with the death domain of a serine/threonine protein kinase known as IRAK; this leads to the autophosphorylation of IRAK^{26,48}. Autophosphorylated IRAK then forms a complex with TRAF6 and this, in turn, results in the oligomerization of TRAF6. At this point the details of the pathway become less clear. Somehow the oligomerization of TRAF6 activates TAK-1, a member of the MAP 3-kinase family⁵², and this leads to the activation of the I κ B kinases. These kinases, in turn, phosphorylate I κ B, leading to its proteolytic degradation and the translocation of NF- κ B to the nucleus^{28,29,46–48}. Concomitantly, members of the activator protein-1 (AP-1) transcription factor family, Jun and Fos^{52–54}, are activated, and both AP-1 transcription factors and NF- κ B are required for cytokine production. Figure 2 also includes other components that have been identified by biochemical and genetic approaches but whose precise location within the pathway has yet to be defined. Although gene deletion studies have supported this activation pathway^{34,55–57}, there is clearly substantial additional complexity. For example, many components of the signalling pathway have homologues: at least three IRAK homologues have been demonstrated, and these are known to compensate partly for each other⁵⁸. Interestingly, although no homologue of MyD88 has yet been found, deletion of this gene does not completely abolish the response to LPS, although the response is substantially delayed⁵⁵. The explanation for the residual responsiveness in currently not known.

Logically, one might assume that downstream events also differ depending on which of the Toll receptors is ligated. The best-studied examples come from *Drosophila*, in which signalling through dToll and 18-wheeler have been shown to produce a distinct pattern of gene products^{19,20}. Much less is known about mammalian systems other than that differential downstream events are linked to signalling through different TLRs. This will be a fruitful area of study in the future, for refined genomic approaches.

The past decade has thus been filled with remarkable advances in our understanding of the mechanisms of innate immune responses to microbial pathogens. In large part the progress was driven by a desire to understand the pathogenesis of a serious clinical syndrome known as septic shock. These advances encompass the identification of membrane receptors for specific products of microbial pathogens, the identification of essential components of intracellular signalling pathways and the characterization of the key mediators produced by cells of the innate immune system in response to infection. This body of information suggests that we will be able further to determine the genetic basis of disease susceptibility and outcome through studies of the genetic variation of essential components. For example, future studies will undoubtedly include a careful analysis of polymorphisms of TLRs. The latter might permit a correlation to be made between polymorphisms and disease outcome in patients in hospital at risk of septic shock. Other important advances will come from a better understanding of how different TLRs interact to recognize microbial pathogens and from a determination of the spatial organization of the various proteins comprising a receptor for a specific class of products released by pathogens. □

- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. A. Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318 (1999).
- Aderem, A. & Underhill, D. M. Mechanisms of phagocytosis in macrophages. *Annu. Rev. Immunol.* **17**, 593–623 (1999).
- Janeway, C. A. Jr & Medzhitov, R. Introduction: the role of innate immunity in the adaptive immune response. *Semin. Immunol.* **10**, 349–350 (1998).
- Ulevitch, R. J. & Tobias, P. S. Recognition of gram-negative bacteria and endotoxin by the innate immune system. *Curr. Opin. Immunol.* **11**, 19–22 (1999).
- Raez, C. R. Biochemistry of endotoxins. *Annu. Rev. Biochem.* **59**, 129–170 (1990).
- Darveau, R. P. Lipid A diversity and the innate host response to bacterial infection. *Curr. Opin. Microbiol.* **1**, 36–42 (1998).
- Lien, E. *et al.* Toll-like receptor 4 imparts ligand-specific recognition of bacterial lipopolysaccharide. *J. Clin. Invest.* **105**, 497–504 (2000).

8. Poltorak, A., Ricciardi-Castagnoli, P., Citterio, S. & Beutler, B. Physical contact between lipopolysaccharide and toll-like receptor 4 revealed by genetic complementation. *Proc. Natl Acad. Sci. USA* **97**, 2163–2167 (2000).
9. Henderson, B., Poole, S. & Wilson, M. Bacterial modulins: a novel class of virulence factors which cause host tissue pathology by inducing cytokine synthesis. *Microbiol. Rev.* **60**, 316–341 (1996).
10. Brennan, P. J. & Nikaido, H. The envelope of mycobacteria. *Annu. Rev. Biochem.* **64**, 29–63 (1995).
11. Kalambaheti, T., Bulach, D. M., Rajakumar, K. & Adler, B. Genetic organization of the lipopolysaccharide O-antigen biosynthetic locus of *Leptospira borgpetersenii* serovar Hardjobovis. *Microb. Pathogen.* **27**, 105–117 (1999).
12. Mitchison, M. *et al.* Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related rfb locus in *Leptospira interrogans* serovar Copenhageni. *J. Bacteriol.* **179**, 1262–1267 (1997).
13. Chaffin, W. L., Lopez-Ribot, J. L., Casanova, M., Gozalbo, D. & Martinez, J. P. Cell wall and secreted proteins of *Candida albicans*: identification, function, and expression. *Microbiol. Mol. Biol. Rev.* **62**, 130–180 (1998).
14. Tachado, S. D. *et al.* Signal transduction in macrophages by glycosylphosphatidylinositols of *Plasmodium*, *Trypanosoma*, and *Leishmania*: activation of protein tyrosine kinases and protein kinase C by inositolglycan and diacylglycerol moieties. *Proc. Natl Acad. Sci. USA* **94**, 4022–4027 (1997).
15. Descoteaux, A., Matlashewski, G. & Turco, S. J. Inhibition of macrophage protein kinase C-mediated protein phosphorylation by *Leishmania donovani* lipophosphoglycan. *J. Immunol.* **149**, 3008–3015 (1992).
16. Ulevitch, R. J. & Tobias, P. S. Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. *Annu. Rev. Immunol.* **13**, 437–457 (1995).
17. Schumann, R. R. *et al.* Structure and function of lipopolysaccharide binding protein. *Science* **249**, 1429–1431 (1990).
18. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* **249**, 1431–1433 (1990).
19. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette *spatzle*/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).
20. Williams, M. J., Rodriguez, A., Kimbrell, D. A. & Eldon, E. D. The 18-wheeler mutation reveals complex antibacterial gene regulation in *Drosophila* host defense. *EMBO J.* **16**, 6120–6130 (1997).
21. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–2088 (1998).
22. Qureshi, S. T. *et al.* Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J. Exp. Med.* **189**, 615–625 (1999). [Erratum, *J. Exp. Med.* **189**, following 1518 (1999)].
23. Hoshino, K. *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J. Immunol.* **162**, 3749–3752 (1999).
24. Imler, J. L. & Hoffmann, J. A. Signaling mechanisms in the antimicrobial host defense of *Drosophila*. *Curr. Opin. Microbiol.* **3**, 16–22 (2000).
25. Anderson, K. V. Toll signaling pathways in the innate immune response. *Curr. Opin. Immunol.* **12**, 13–19 (2000).
26. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
27. Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
28. Yang, R. B. *et al.* Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* **395**, 284–288 (1998).
29. Kirschning, C. J., Wesche, H., Merrill, A. T. & Rothe, M. Human toll-like receptor 2 confers responsiveness to bacterial lipopolysaccharide. *J. Exp. Med.* **188**, 2091–2097 (1998).
30. Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
31. Heine, H. *et al.* Cutting edge: cells that carry A null allele for toll-like receptor 2 are capable of responding to endotoxin. *J. Immunol.* **162**, 6971–6975 (1999).
32. Shimazu, R. *et al.* MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med.* **189**, 1777–1782 (1999).
33. Flo, T. H. *et al.* Human toll-like receptor 2 mediates monocyte activation by *Listeria monocytogenes*, but not by group B streptococci or lipopolysaccharide. *J. Immunol.* **164**, 2064–2069 (2000).
34. Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
35. Lien, E. *et al.* Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *J. Biol. Chem.* **274**, 33419–33425 (1999).
36. Takeuchi, O. *et al.* Cutting edge: preferentially the R-stereoisomer of the mycoplasmal lipopeptide macrophage-activating lipopeptide-2 activates immune cells through a toll-like receptor. *J. Immunol.* **164**, 554–557 (2000).
37. Hirschfeld, M. *et al.* Cutting edge: inflammatory signaling by *Borrelia burgdorferi* lipoproteins is mediated by toll-like receptor 2. *J. Immunol.* **163**, 2382–2386 (1999).
38. Ohashi, K., Burkart, V., Flohe, S. & Kolb, H. Cutting edge: heat shock protein 60 is a putative endogenous ligand of the toll-like receptor-4 complex. *J. Immunol.* **164**, 558–561 (2000).
39. Means, T. K. *et al.* The CD14 ligands lipoarabinomannan and lipopolysaccharide differ in their requirement for Toll-like receptors. *J. Immunol.* **163**, 6748–6755 (1999).
40. Means, T. K. *et al.* Human toll-like receptors mediate cellular activation by *Mycobacterium tuberculosis*. *J. Immunol.* **163**, 3920–3927 (1999).
41. Brightbill, H. D. *et al.* Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. *Science* **285**, 732–736 (1999).
42. Underhill, D. M., Ozinsky, A., Smith, K. D. & Aderem, A. Toll-like receptor-2 mediates mycobacteria-induced proinflammatory signaling in macrophages. *Proc. Natl Acad. Sci. USA* **96**, 14459–14463 (1999).
43. Chaudhary, P. M. *et al.* Cloning and characterization of two Toll/Interleukin-1 receptor-like genes TIL3 and TIL4: evidence for a multi-gene receptor family in humans. *Blood* **91**, 4020–4027 (1998).
44. Takeuchi, O. *et al.* TLR6: a novel member of an expanding toll-like receptor family. *Gene* **231**, 59–65 (1999).
45. Levaschina, E. A. *et al.* Constitutive activation of toll-mediated antifungal defense in serpin-deficient *Drosophila*. *Science* **285**, 1917–1919 (1999).
46. Medzhitov, R. *et al.* MyD88 is an adaptor protein in the hToll/IL-1 receptor family signaling pathways. *Mol. Cell* **2**, 253–258 (1998).
47. Yang, R. B., Mark, M. R., Gurney, A. L. & Godowski, P. J. Signaling events induced by lipopolysaccharide-activated toll-like receptor 2. *J. Immunol.* **163**, 639–643 (1999).
48. Muzio, M., Natoli, G., Saccani, S., Levrero, M. & Mantovani, A. The human toll signaling pathway: divergence of nuclear factor κ B and JNK/SAPK activation upstream of tumor necrosis factor receptor-associated factor 6 (TRAF6). *J. Exp. Med.* **187**, 2097–2101 (1998).
49. Medzhitov, R. & Janeway, C. Jr Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* **173**, 89–97 (2000).
50. Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* **7**, 837–847 (1997).
51. Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. TRAF6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
52. Ninomiya-Tsuji, J. *et al.* The kinase TAK1 can activate the NIK- κ B as well as the MAP kinase cascade in the IL-1 signalling pathway. *Nature* **398**, 252–256 (1999).
53. Kopp, E. *et al.* ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway. *Genes Dev.* **13**, 2059–2071 (1999).
54. Irie, T., Muta, T. & Takeshige, K. TAK1 mediates an activation signal from toll-like receptor(s) to nuclear factor- κ B in lipopolysaccharide-stimulated macrophages. *FEBS Lett.* **467**, 160–164 (2000).
55. Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
56. Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
57. Thomas, J. A. *et al.* Impaired cytokine signaling in mice lacking the IL-1 receptor-associated kinase. *J. Immunol.* **163**, 978–984 (1999).
58. Wesche, H. *et al.* IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family. *J. Biol. Chem.* **274**, 19403–19410 (1999).

CD1-restricted T-cell responses and microbial infection

Se-Ho Park & Albert Bendelac

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

CD1, a conserved family of major histocompatibility (MHC)-like glycoproteins in mammals, specializes in capturing lipid rather than peptide antigen for presentation to T lymphocytes. The principles and mechanisms of this newly discovered immune strategy differ markedly from those governing classical MHC-peptide presentation. They might be exploited for the design of new lipid-based microbial vaccines and adjuvants.

A key feature of the T lymphocytes of the adaptive immune system is their ability to detect minute concentrations of pathogen-derived peptides presented by MHC molecules on the cell surface. MHC molecules have a peptide-binding groove that anchors short peptides on the basis of the presence of a motif consisting of two to four residues. They have been under strong evolutionary pressure to diversify their peptide-binding groove, probably to meet the challenges of changing microbial environments, and are consequently highly polymorphic.

In contrast with MHC, with which it shared a common ancestor around or before the mammalian radiation, CD1 is a family of non-polymorphic genes¹ that seems to have evolved to present lipid and glycolipid antigens to T cells². It comprises up to five distinct genes (isotypes) that are conserved in different mammalian species and can be separated into two groups on the basis of sequence homology (Fig. 1). Group 1, which comprises CD1a, CD1b, CD1c and CD1e, is present in humans but absent from mouse and rat. Group 2, which includes only CD1d, is found in all species studied so far. CD1 genes encode heavy chains that are only 30% identical to that of MHC class I, yet associate with β_2 -microglobulin to adopt a very similar fold. Since its discovery, CD1 has attracted much attention but knowledge of its possible functions has been acquired only slowly. In retrospect, this was probably because, unlike MHC, CD1 is not polymorphic and thus did not seem to be an immune-response gene in genetic studies. In addition, several CD1 genes are poorly expressed in uninduced peripheral blood cells, the common source of antigen-presenting cells used for studies *in vitro*.

CD1 was recently shown to present microbial and self lipid antigens to T cells. There is now evidence for both an innate and an adaptive pathway of lipid antigen recognition, opening new avenues for studying the evolution of host-pathogen relations and for the design of novel vaccines and adjuvants.

Lipid recognition by CD1-restricted T lymphocytes

Lipids that are structural components of the mycobacterial cell wall such as mycolic acids³, glucose monomycolate⁴ and derivatives of lipoarabinomannans (LAM) such as phosphatidylinositolmannosides⁵ are presented by CD1b, whereas hexose-1-phosphoisoprenoids⁶ are presented by CD1c, and these antigens seem to be dominant because they are found to be recognized by $\alpha\beta$ -T-cell-receptor (TCR)-expressing cell lines from different individuals. These anti-

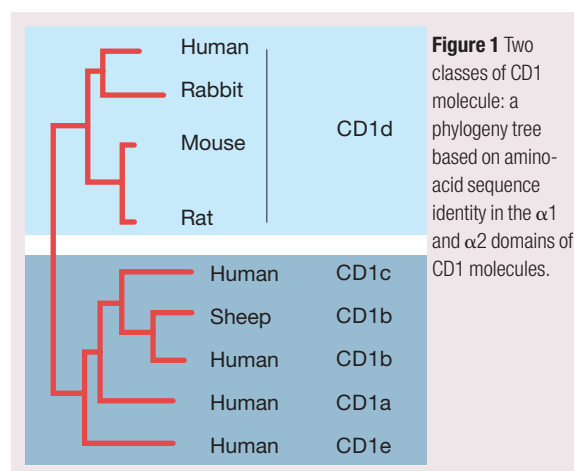


Figure 1 Two classes of CD1 molecule: a phylogeny tree based on amino acid sequence identity in the $\alpha 1$ and $\alpha 2$ domains of CD1 molecules.

gens are particularly interesting because they represent a large biosynthetic investment by mycobacteria, to which they confer a uniquely resistant cell wall, and are absent, or exist in modified form, in vertebrates⁷ (Fig. 2). Other mycobacterial lipids presented by CD1a, CD1b or CD1c as well as microbial lipids of non-mycobacterial origin have yet to be characterized.

α -Galactosylceramide (α -GalCer), a glycolipid originally extracted from marine sponges⁸, can be recognized in picomolar concentrations by an entire population of mouse and human CD1d-restricted lymphocytes that express a semi-invariant TCR and exert potent effector and regulatory functions⁹.

Self ceramides such as gangliosides can also be presented by CD1b (ref. 10). Because these antigens are highly expressed in brain tissues, recognition might regulate or contribute to autoimmune damage in diseases such as multiple sclerosis.

Thus, although hydrophobic peptides might also be presented by CD1 (ref. 11), the expanding list of lipids recognized by CD1-restricted T cells suggests that, in contrast with MHC, CD1 evolved to present lipids rather than peptides.

CD1: lipid-presenting molecules with broad specificity

Although the way in which CD1 molecules bind lipids has not yet been solved, the crystal structure of mouse CD1d revealed a unique feature that probably mediates this property¹². CD1d has a deep ligand-binding groove, about 30 Å long by 10–15 Å wide, made of two large electrostatically

neutral pockets lined with clustered hydrophobic residues that create continuous regions of hydrophobicity, seemingly well suited to bind fatty-acid chains up to 30 carbons long (Fig. 3). Such binding would seem to be rather non-specific, allowing many types of lipids to be buried in many different positions in the groove. As functional studies suggested that the CD1-restricted cells could discriminate between lipid antigens with different hydrophilic caps but not with different fatty-acid chains^{4,9}, the current model suggests that the polar head group is important both for the positioning and for the presentation of antigenic residues to the TCR¹³. Although exquisite specificity can be imparted by a single carbohydrate residue⁴, it is predicted, given the expected size of the TCR/CD1 contact area, that up to three or four proximal carbohydrates, and their various modes of branching, might be critical for interactions with the TCR.

The conservation of CD1 isotypes across species, that is, the fact that human CD1d is closest to mouse CD1d (~70% identity) than to human CD1b (~50% identity) for example, suggests that they have evolved to perform specialized functions. One possibility is that each isotype binds a different spectrum of lipids. However, functional studies suggest that there must be a large overlap in the binding specificities of at least some of the CD1 isotypes. For example, ceramide-based and phosphatidyl-based foreign lipids as well as self lipids are bound by the divergent CD1b (group 1) and CD1d (group 2) isotypes; there is strong evidence that the two fatty-acid chains of these lipids are required for optimal binding^{5,10,14,15}. The binding properties of CD1c might be different, as the first mycobacterial antigen associated with CD1c has a single hydrocarbon chain⁶. The binding constants of some lipids to CD1 in solution have been estimated, but their interpretation is complicated by the tendency of several lipids to form micelles and by the possibility that, physiologically, binding might not occur in solution but on transfer from cellular membranes in specialized compartments. Nevertheless, K_d values in the range 10^{-7} M for CD1d–phosphatidylinositol, CD1b–LAM and CD1b–phosphatidylinositolmannoside interactions have been suggested^{16,17}. Perhaps more significant is the finding that CD1 can bind lipids stably, with a half-life of various CD1–lipid complexes ranging from half an hour to more than 12 hours in a manner reminiscent of peptide trapping by MHC class II^{14,15,17}. Further studies are required to explain how different lipids bind to different CD1 isotypes. One especially challenging question relates to the competition faced by foreign microbial lipids with the extremely abundant endogenous lipids that can potentially bind CD1. Although cellular glycosylphosphatidylinositols (GPIs) were suggested to be the main endogenous ligand of CD1d (ref. 16), evidence presented so far is not conclusive and the main types of lipid naturally bound to CD1 remain to be identified.

Assembly and transport of CD1– β_2 -microglobulin complexes occur independently of transporters required for MHC molecules. Lipid binding might occur in the secretory pathway, directly at the cell surface¹⁸ or only after internalization in an acidified compartment, which has been suggested to favour the opening of the CD1 groove and is speculated to contain enzymes capable of trimming the polysaccharide or the fatty-acid components of glycolipids¹⁷. Although the details of CD1 trafficking are not yet known, the steady-state intracellular locations of the different isotypes are remarkably different, covering different compartments of the endocytic pathway^{19,20}. CD1b and CD1d are concentrated mainly in the late endosome or lysosome, CD1c in the late endosome and CD1a in the early or recycling endosome. Access to the endocytic pathway is regulated by a tyrosine-based motif in the cytoplasmic tail of CD1 that differs between CD1b, CD1c and CD1d. Mutant CD1b and CD1d molecules lacking this motif lose their ability to present at least some antigens to T cells, indicating that they sample this compartment before reaching the cell surface^{21,22}. Integrating these findings with the notion that lipids with different structural properties travel to or are structural components of distinct endocytic compartments, a persuasive model suggests that CD1

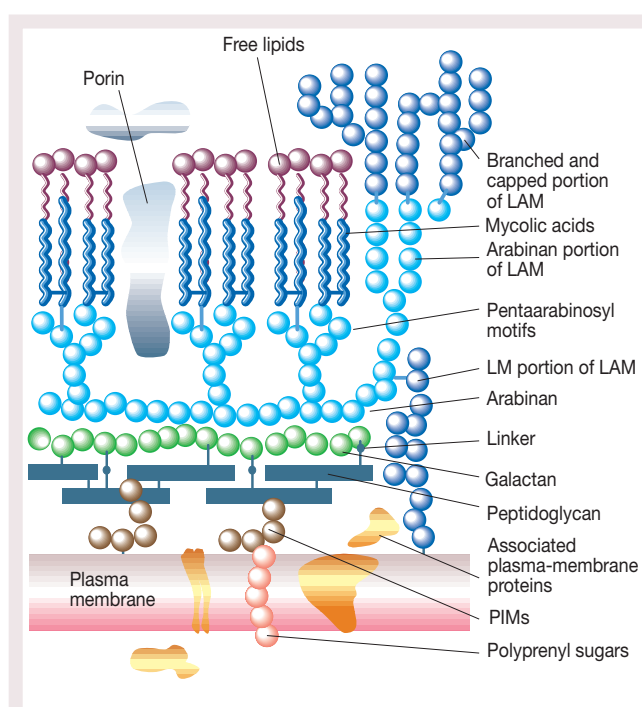


Figure 2 The *Mycobacterium tuberculosis* cell envelope. (Modified from ref. 7, with permission.) LAM, lipoarabinomannans; LM, lipomannans; PIM, phosphatidylinositol mannosides.

isotypes evolved to target distinct lipids in distinct cellular compartments^{19,23} (Fig. 4).

An adaptive pathway of microbial lipid recognition

The prevalence of CD1-restricted responses *in vivo* during infection and their protective value have been difficult to assess until recently. Initial reports were based on a limited number of T-cell lines derived *in vitro* by repeated stimulation of PBLs with CD1-expressing cells and mycobacterial extracts. They involved CD4⁺8⁺ T cells, a rare cell type. More recently, cell lines were derived that expressed CD8, a generic co-receptor for MHC class I and a putative co-receptor for CD1 that is used by a large fraction of lymphocytes². CD4⁺ cell lines have also been reported²⁴, although CD4 does not seem to function as a co-receptor for CD1. Importantly, Moody *et al.*⁶ demonstrated a significant CD1c-restricted, T-cell proliferative response to mannosyl phosphodolichols, semi-synthetic analogues of mycobacterial hexose-1-phosphoisoprenoids, upon *in vitro* stimulation of peripheral blood lymphocytes from patients recently infected with *Mycobacterium tuberculosis*. This observation is a major step in supporting the notion that immunization against mycobacterial lipids does occur *in vivo* during natural infections.

The protective value of these CD1-restricted antimicrobial responses remains to be evaluated. CD1 isotypes are expressed selectively by antigen-presenting cells such as dendritic cells, macrophages and subsets of B cells, but apart from CD1d expression in hepatocytes they are generally not expressed in solid tissues^{2,25}. The only CD1 isotypes so far found to present microbial lipid antigens are CD1a, CD1b and CD1c, which are not present in mice or rats, precluding the use of these animal models. CD1d, which does exist in rodents and, like CD1b, binds LAM¹⁵, has nevertheless not been associated with antimycobacterial defence. Thus, the key infection and vaccination experiments will have to be performed in non-human primates or in guinea-pigs, which express the CD1a, CD1b and CD1c isotypes and are susceptible to infection with *Mycobacterium tuberculosis*. There are some indications that CD1-restricted cells might be useful in infection. Indeed, experiments *in vitro* showed that human

CD1-restricted cells could display various effector properties against cells infected with live mycobacteria. CD8⁺, CD4⁺ and CD4⁺8⁺ cell lines released interferon- γ , an important anti-mycobacterial cytokine. CD8 also used perforin and granulysin stored in cytolytic granules to induce apoptosis in infected macrophages and kill mycobacteria, respectively, whereas CD4⁺8⁺ cells induced apoptosis through the Fas pathway²⁶.

Innate functions of CD1-restricted cells

Selection and expansion of a few antigen-specific cells from a large pool of naive B and T lymphocytes is a key feature of adaptive immune responses. In contrast, some peculiar subsets of lymphocytes are considerably expanded in all adults (and rare in newborns), presenting a 'natural memory' phenotype indicative of chronic stimulation, despite an apparent lack of exposure to foreign antigen. These 'natural memory' cells express germline-encoded semi-invariant antigen receptors that often display latent or overt reactivity to self antigens. Because collectively they represent up to 10–50% of the adult lymphocyte population in various tissues, including the peritoneum, gut, liver, spleen, bone marrow and blood, they can display powerful functions at the onset of immune responses, long before classical B or T lymphocytes have been recruited in significant numbers. Functionally, therefore, they bridge the innate and the adaptive immune response. Examples of such 'natural memory' cells include B-1 B cells expressing the phosphatidylcholine specificity encoded by VH11-V κ 9 or VH12-V κ 4, which recognize altered cellular membranes²⁷, or those with anti-phosphorylcholine specificity, which are vital for the innate recognition and clearance of microbial elements²⁸. Another example of an innate lymphocyte that is also deficient before adulthood is the natural killer cell, which is essential for protection against herpes viruses²⁹. Unique among MHC and MHC-like molecules, CD1 is recognized by the antigen receptors of at least two prominent innate cell subsets, the CD1d-restricted mouse V α 14 and human V α 24 natural killer T cells³⁰ and the CD1c-specific human V δ 1 $\gamma\delta$ T cells³¹.

CD1d-restricted natural killer T cells

The main T-cell subset associated with CD1d is the natural killer T cell, which expresses a TCR made of an invariant, germline-encoded, TCR α chain (mouse V α 14-J α 281 or human V α 24-J α Q) associated with diverse mouse V β 8 or human V β 11 chains, which is autoreactive to CD1d (refs 25, 32). This TCR recognizes a conserved family of glycolipids. It has been suggested that self and foreign GPIs are recognized by natural killer T cells, ensuring strong responses to GPI-associated antigens such as malarial circumsporozoite³³; although the model is attractive, the data are controversial³⁴. In contrast, it has been well established that virtually all mouse and human natural killer T cells can recognize α -GalCer, a glycolipid originally extracted from marine sponges. Although individual natural killer T-cell TCRs might also recognize other self lipid antigens such as phosphatidylinositol or phosphatidylglycerol³⁵, the generic recognition of a picomolar concentration of α -GalCer by an entire cell population expressing a restricted family of autoreactive antigen receptors¹⁵ reinforces the parallel with the autoreactive VH11-V κ 9 or VH12-V κ 4 phosphatidylcholine-specific B-1 B cells. Although the mammalian counterpart of α -GalCer has yet to be identified, it is conceivable that it is induced under a range of pathological and inflammatory conditions to activate natural killer T cells.

The activation of natural killer T cells *in vivo* induces a prompt series of cellular activation events that is highly reminiscent of other adjuvant-induced reactions because it leads to the activation of innate cells such as natural killer cells and dendritic cells, the activation of adaptive cells such as B cells and T cells, the induction of co-stimulatory molecules and the abrupt release of cytokines such as interleukin-4 and interferon- γ ^{36–38}. In addition, the activated natural killer T cells can themselves bring about killing mediated by Fas and perforin. The full activation cascade can be recruited by the

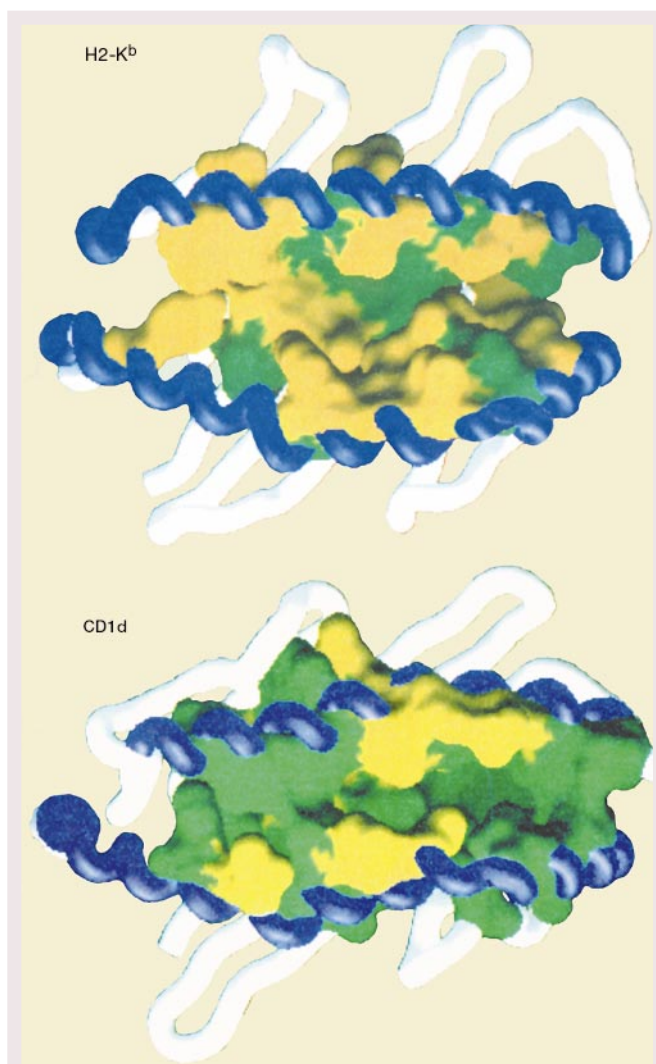


Figure 3 A comparison of the hydrophobicity of CD1d and H2-K^b antigen-binding grooves. The entire binding groove is represented as a surface, with hydrophobic residues shown in green and non-hydrophobic residues in yellow. The portions of the α -helices that do not contribute to the binding groove are shown in blue and the remainder of the molecule is white. Note the continuous hydrophobic lining of the CD1 groove compared with that in H2-K^b. The figure was generated with the GRASP software from published coordinates for CD1d and H2-K^b.

engagement of natural killer T TCR, for example after the injection of α -GalCer *in vivo*. Alternatively, powerful T-helper-cell type 1 (T_H1) functions can be selectively triggered by cytokines such as interleukin-12 released by infected macrophages or dendritic cells. These impressive functions are likely to be correlated with the important role of natural killer T cells in conditions such as autoimmune diabetes^{39–41}, rejection of established tumours or the prevention of chemically induced tumours⁴². Natural killer T cells are thought to contribute to antimicrobial immunity through their capacity to influence the T_H1–T_H2 polarization^{43,44}. These functions are currently being actively investigated, because it seems obvious that, whether for the apparent benefit of the host or the pathogen⁴⁵, such a powerful adjuvant pathway is recruited during at least some microbial infections.

CD1c-reactive cells

Interestingly, in humans, a group 1 CD1 isotype is also associated with innate immunity. CD1c is recognized by a major subset of cytolytic $\gamma\delta$ T cells, especially in the gut, that express V δ 1 $\gamma\delta$ TCRs of limited

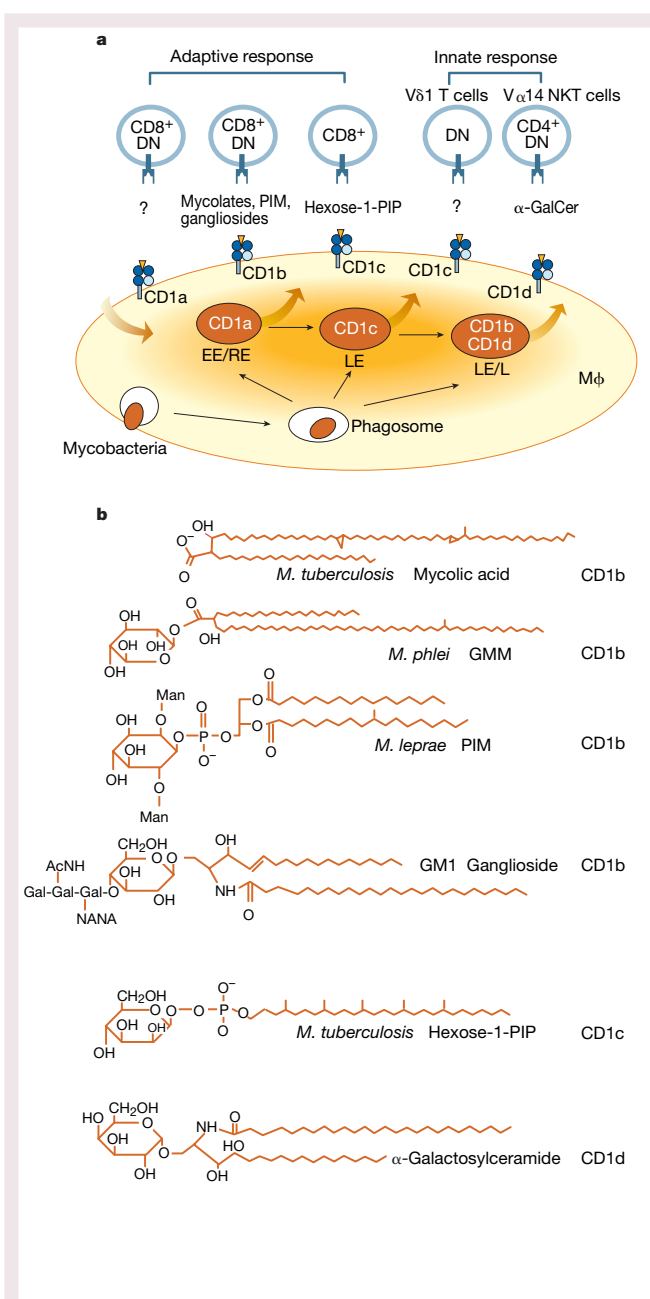


Figure 4 Cellular location of CD1a, CD1b, CD1c, CD1d and corresponding CD1-restricted T-cell subsets and lipid antigens. **a**, CD1 isoforms are thought to sample distinct compartments of the endocytic pathway before reaching the cell surface to present different families of lipids, of self or foreign (for example mycobacterial lipids exported from the phagosome) origins. Distinct innate and adaptive T-cell subsets recognize these antigens. Mφ, macrophage. **b**, Lipid structures and their corresponding CD1-presenting molecules. DN, double negative (CD4[−]8[−]); GMM, glucose monomycolate; PIM, phosphatidylinositol mannoside; PIP, phosphoisoprenoid; EE/RE, early/recycling endosome; LE, late endosome; LE/L, late endosome/lysosome.

germline diversity and whose sequences suggest antigen-driven selection⁴⁶. As with CD1d-autoreactive natural killer T cells, the self antigens involved have not yet been identified³¹, but it seems likely that their induction serves a similar purpose in the prompt triggering of innate immune cascades in response to various stresses and infections.

Natural-killer-like functions of CD1-restricted innate subsets

In addition to their TCR, the CD1d-restricted natural killer T cells and the CD1c-restricted Vδ1 T cells conspicuously express a battery

of receptors and functions that are normally expressed by natural killer cells, further linking them to innate immunity. This enables them to be activated or regulated by several of the same signals that trigger natural killer cells, exerting natural-killer-like functions that, in retrospect, might have been improperly attributed exclusively to classical natural killer cells.

Developing novel, lipid-based adjuvant and vaccines

The recognition that CD1, an orphan family of MHC-like molecules, has an entirely new role in antigen presentation to T lymphocytes came as a surprise in a closely studied area of immunology, raising the prospect of novel, lipid-based vaccines. Considerable work will be needed to clarify the mode of lipid-antigen presentation and the corresponding repertoire of TCR specificities and cellular functions. However, evidence is accumulating that indicates that the CD1 pathway of antigen presentation also differs from the MHC pathway with regard to the context within which it operates. Indeed, although microbial lipid antigens, because they are conserved, might seem an ideal target for T-cell immune recognition and vaccination, the identification of the first lipid antigens recognized by CD1-restricted T cells, such as mycolic acids and LAM, reminds us that these antigens — as well as many other microbial lipids, glycolipids and lipoproteins — are already targeted by a variety of powerful, but poorly characterized, non-T-cell-mediated innate recognition pathways^{47,48}. The innate pathways are collectively responsible for the ‘adjuvant effect’, the black box of modern immunology, which results in the induction of the co-stimulatory molecules and cytokines required for adaptive responses (see the review in this issue by Aderem and Ulevitch, pages 782–787). Furthermore, prominent populations of CD1d-restricted cells such as natural killer T cells themselves exert these powerful adjuvant effects and can profoundly influence the outcome of infectious and non-infectious diseases. This intricate relationship between diffuse, innate mechanisms of defence and adaptive cellular phenomena firmly anchors CD1 at the interface between innate and adaptive immunity. It offers an exciting but difficult challenge to future studies that aim at dissecting these effects *in vivo* as well as a promising avenue of future research with the potential to influence profoundly the rational design of lipid-based vaccines and adjuvants. □

- Calabi, F., Jarvis, J. M., Martin, L. H. & Milstein, C. Two classes of CD1 genes. *Eur. J. Immunol.* **19**, 285–292 (1989).
- Porcelli, S. A. & Modlin, R. L. The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu. Rev. Immunol.* **17**, 297–329 (1999).
- Beckman, E. M. *et al.* Recognition of a lipid antigen by CD1-restricted αβ T cells. *Nature* **372**, 691–694 (1994).
- Moody, D. B. *et al.* Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* **278**, 283–286 (1997).
- Sieling, P. A. *et al.* CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science* **269**, 227–230 (1995).
- Moody, D. B. *et al.* CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection. *Nature* **404**, 884–888 (2000).
- Lee, R. E., Brennan, P. J. & Besra, G. S. *Mycobacterium tuberculosis* cell envelope. *Curr. Top. Microbiol. Immunol.* **215**, 1–27 (1996).
- Natori, T., Morita, M., Akimoto, K. & Koezuka, Y. Agelasphins, novel antitumor and immunostimulatory cerebroside from the sponge *Agelas mauritanicus*. *Tetrahedron* **50**, 2771–2784 (1994).
- Kawano, T. *et al.* CD1d-restricted and TCR-mediated activation of Vα14 NKT cells by glycosylceramides. *Science* **278**, 1626–1629 (1997).
- Shamshiev, A. *et al.* Self glycolipids as T-cell autoantigens. *Eur. J. Immunol.* **29**, 1667–1675 (1999).
- Castano, A. R. *et al.* Peptide binding and presentation by mouse CD1. *Science* **269**, 223–226 (1995).
- Zeng, Z. H. *et al.* The crystal structure of murine CD1: an MHC-like fold with a large hydrophobic antigen binding groove. *Science* **277**, 339–345 (1997).
- Porcelli, S. A. & Brenner, M. B. Antigen presentation: mixing oil and water. *Curr. Biol.* **7**, R508–R511 (1997).
- Naidenko, O. V. *et al.* Binding and antigen presentation of ceramide-containing glycolipids by soluble mouse and human CD1d molecules. *J. Exp. Med.* **190**, 1069–1080 (1999).
- Benlagha, K., Weiss, A., Beavis, A., Teyton, L. & Bendelac, A. *In vivo* identification of glycolipid antigen specific T cells using fluorescent CD1d tetramers. *J. Exp. Med.* **191**, 1895–1903 (2000).
- Joyce, S. *et al.* Natural ligand of mouse CD1d1: cellular glycosylphosphatidylinositol. *Science* **279**, 1541–1544 (1998).
- Ernst, W. A. *et al.* Molecular interaction of CD1b with lipoglycan antigens. *Immunity* **8**, 331–340 (1998).
- Burdin, N. *et al.* Selective ability of mouse CD1 to present glycolipids: α-galactosylceramide specifically stimulates Vα14 NKT lymphocytes. *J. Immunol.* **161**, 3271–3281 (1998).

19. Sugita, M. *et al.* Separate pathways for antigen presentation by CD1 molecules. *Immunity* **11**, 743–752 (1999).
20. Schaible, U. E., Hagens, K., Fischer, K., Collins, H. L. & Kaufmann, S. H. Intersection of group I CD1 molecules and mycobacteria in different intracellular compartments of dendritic cells. *J. Immunol.* **164**, 4843–4852 (2000).
21. Jackman, R. M. *et al.* The tyrosine-containing cytoplasmic tail of CD1b is essential for its efficient presentation of bacterial lipid antigens. *Immunity* **8**, 341–351 (1998).
22. Chiu, Y. H. *et al.* Distinct subsets of CD1d-restricted T cells recognize self-antigens loaded in different cellular compartments. *J. Exp. Med.* **189**, 103–110 (1999).
23. Briken, V., Jackman, R. M., Watts, G. F. M., Rogers, R. A. & Porcelli, S. A. Human CD1b and CD1c isoforms survey different intracellular compartments for the presentation of microbial lipid antigens. *J. Exp. Med.* **192**, 281–288 (2000).
24. Sieling, P. A. *et al.* Evidence for human CD4⁺ T cells in the CD1-restricted repertoire: derivation of mycobacteria-reactive T cells from leprosy lesions. *J. Immunol.* **164**, 4790–4796 (2000).
25. Bendelac, A., Rivera, M. N., Park, S.-H. & Roark, J. H. Mouse CD1-specific NK1 T cells. Development, specificity, and function. *Annu. Rev. Immunol.* **15**, 535–562 (1997).
26. Stenger, S. *et al.* Differential effects of cytolytic T cell subsets on intracellular infection. *Science* **276**, 1684–1687 (1997).
27. Tatu, C., Ye, J., Arnold, L. W. & Clarke, S. H. Selection at multiple checkpoints focuses V(H)12 B cell differentiation toward a single B-1 cell specificity. *J. Exp. Med.* **190**, 903–914 (1999).
28. Benedict, C. L. & Kearney, J. F. Increased junctional diversity in fetal B cells results in a loss of protective anti-phosphorylcholine antibodies in adult mice. *Immunity* **10**, 607–617 (1999).
29. Biron, C. A., Nguyen, K. B., Pien, G. C., Cousens, L. P. & Salazar-Mather, P. T. Natural killer cells in antiviral defense: function and regulation by innate cytokines. *Annu. Rev. Immunol.* **17**, 189–220 (1999).
30. Bendelac, A. *et al.* CD1 recognition by mouse NK1⁺ T lymphocytes. *Science* **268**, 863–865 (1995).
31. Spada, F. M. *et al.* Self-recognition of CD1 by γ/δ T cells: implications for innate immunity. *J. Exp. Med.* **191**, 937–948 (2000).
32. Park, S. H., Benlagha, K., Lee, D., Balish, E. & Bendelac, A. Unaltered phenotype, tissue distribution and function of V α 14(+) NKT cells in germ-free mice. *Eur. J. Immunol.* **30**, 620–625 (2000).
33. Schofield, L. *et al.* CD1d-restricted immunoglobulin G formation to GPI-anchored antigens mediated by NKT cells. *Science* **283**, 225–229 (1999).
34. Molano, A. *et al.* Cutting edge. The IgG response to the circumsporozoite protein is MHC class II-dependent and CD1d-independent: exploring the role of glycosylphosphatidylinositols in NKT cell activation and anti-malarial responses. *J. Immunol.* **164**, 5005–5009 (2000).
35. Gumperz, J. E. *et al.* Murine CD1d-restricted T cell recognition of cellular lipids. *Immunity* **12**, 211–221 (2000).
36. Burdin, N., Brossay, L. & Kronenberg, M. Immunization with α -galactosylceramide polarizes CD1-reactive NK T cells towards Th2 cytokine synthesis. *Eur. J. Immunol.* **29**, 2014–2025 (1999).
37. Carnaud, C. *et al.* Cross-talk between cells of the innate immune system: NKT cells rapidly activate NK cells. *J. Immunol.* **163**, 4647–4650 (1999).
38. Kitamura, H. *et al.* The natural killer T (NKT) cell ligand α -galactosylceramide demonstrates its immunopotentiating effect by inducing interleukin (IL)-12 production by dendritic cells and IL-12 receptor expression on NKT cells. *J. Exp. Med.* **189**, 1121–1128 (1999).
39. Hammond, K. J. L. *et al.* α/β -T cell receptor (TCR) CD4⁺CD8⁺ (NKT) thymocytes prevent insulin-dependent diabetes mellitus in nonobese diabetic (NOD)/Lt mice by the influence of interleukin (IL)-4 and/or IL-10. *J. Exp. Med.* **187**, 1047–1056 (1998).
40. Lehuen, A. *et al.* Overexpression of natural killer T cells protects V α 14-J α 281 transgenic nonobese diabetic mice against diabetes. *J. Exp. Med.* **188**, 1831–1839 (1998).
41. Wilson, S. B. *et al.* Extreme Th1 bias of invariant V α 24J α Q T cells in type I diabetes. *Nature* **391**, 177–181 (1998).
42. Smyth, M. J. *et al.* Differential tumor surveillance by natural killer (NK) and NKT cells. *J. Exp. Med.* **191**, 661–668 (2000).
43. Cui, J. *et al.* Inhibition of T helper cell type 2 cell differentiation and immunoglobulin E response by ligand-activated V α 14 natural killer T cells. *J. Exp. Med.* **190**, 783–792 (1999).
44. Singh, N. *et al.* Cutting edge: activation of NK T cells by CD1d and α -galactosylceramide directs conventional T cells to the acquisition of a Th2 phenotype. *J. Immunol.* **163**, 2373–2377 (1999).
45. Shinkai, K. & Locksley, R. M. CD1, tuberculosis, and the evolution of major histocompatibility complex molecules. *J. Exp. Med.* **191**, 907–914 (2000).
46. Beldjord, K., Beldjord, C., Macintyre, E., Even, P. & Sigaux, F. Peripheral selection of V δ 1+ cells with restricted T cell receptor delta gene junctional repertoire in the peripheral blood of healthy donors. *J. Exp. Med.* **178**, 121–127 (1993).
47. Besra, G. & Chatterjee, D. in *Tuberculosis: Pathogenesis, Protection and Control*. (ed. Bloom, B.) 285–306 (American Society for Microbiology, Washington, DC, 1994).
48. Almeida, I. C. *et al.* Highly purified glycosylphosphatidylinositols from *Trypanosoma cruzi* are potent proinflammatory agents. *EMBO J.* **19**, 1476–1485 (2000).

Acknowledgements

We thank S. Balk, M. Bonneville, M. Ferguson, J.-J. Fournie, S. Joyce, S. Kaufmann, Y. Koezuka, H. R. MacDonald, I. Orme, S. Porcelli, L. Schofield, L. Teyton, C.-R. Wang and numerous other colleagues for helpful discussions and for sharing unpublished results; P. Brennan for permission to reproduce his model of the mycobacterial cell wall; members of our laboratory for stimulating discussions; H. Piper for help in the preparation of Fig. 2; and B. Jabri and P. Matzinger for critical review of the manuscript. This work was supported by grants from the NIH, American Cancer Society and Juvenile Diabetes Foundation.

Vaccines against intracellular infections requiring cellular immunity

Robert A. Seder* & Adrian V. S. Hill†

*Clinical Immunology Section, Laboratory of Clinical Investigation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

†Molecular Immunology Group, Institute of Molecular Medicine, Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

Vaccines against a variety of infectious diseases represent one of the great triumphs of medicine. The immune correlates of protection induced by most current vaccines seem to be mediated by long-lived humoral immune responses. By contrast, there are no currently available vaccines that are uniformly effective for diseases such as HIV, malaria and tuberculosis, in which the cellular immune response might be crucial in mediating protection. Here we examine the mechanisms by which long-lived cellular immune responses are generated and maintained *in vivo*. We then discuss current approaches for vaccination against diseases in which cellular immune responses are important for protection.

The origins of acquired immunity against infectious disease were observed more than 2,000 years ago. In 430 BC, Thucydides described the plague of Athens in his historical account of the Peloponnesian war:

It was with those who had recovered from the disease that the sick and the dying found most compassion. These knew what it was from experience, and had now no fear for themselves; for the same man was never attacked twice — never at least fatally.

At the end of the eighteenth century, Jenner's observation that vaccination with cowpox virus protected individuals from smallpox provided the scientific basis for the concept of vaccination. Today, vaccines for a variety of bacterial and viral diseases have had a major impact worldwide in reducing both morbidity and mortality. The main reason for the success of current vaccines is that, for most viral and bacterial infections, primary protection is thought to be mediated by a long-lived humoral immune response through the production of antibodies. It is noteworthy that such responses are readily induced by many different vaccine formulations. By contrast, for infections such as tuberculosis, malaria and HIV, which now bring about a substantial proportion of the deaths worldwide from infection, there are no uniformly effective vaccines. Furthermore, although humoral immunity could be important in preventing infection by HIV and certain stages of malaria infection, it is the cellular immune response that is central to the mediation of protection in all of these diseases. The focus of this review is our understanding of how long-lived cellular immune responses can be induced *in vivo* after vaccination.

Acquired cellular immunity

The acquired cellular immune response is composed of CD4⁺ and CD8⁺ T cells. CD4⁺ T cells recognize proteins — commonly referred to as antigens — after they have been processed by a specific group of cells termed antigen-presenting cells. CD4⁺ T cells recognize antigens that have been processed through the exogenous pathway by antigen-

presenting cells (such as dendritic cells, macrophages and B cells) expressing major histocompatibility complex (MHC) class II molecules (Fig. 1). After the recognition of antigen in association with MHC class II molecules, CD4⁺ T cells become activated, and differentiation can occur into functional subsets termed T helper 1 (T_H1)-type and T helper 2 (T_H2)-type cells. The distinction of these subsets is based on their production of a variety of proteins called cytokines. The signature cytokines for T_H1 and T_H2 cells are interferon (IFN)- γ and interleukin (IL)-4, respectively. T_H1 cells, through their production of IFN- γ , mediate the killing of organisms responsible for a variety of intracellular infections. For many intracellular infections, the induction of a functional T_H1 response is crucially dependent on another cytokine, IL-12, which is produced by antigen-presenting cells such as macrophages or dendritic cells after exposure to the pathogen at the initiation of the immune response. Thus, in response to many intracellular infections, IL-12 is the inducer cytokine of T_H1 cells, and IFN- γ is the effector cytokine that mediates protection. CD8⁺ T cells recognize antigens that are processed and presented by cells expressing MHC class I molecules. CD8⁺ T cells also mediate their effector function through the production of cytokines such as IFN- γ and tumour necrosis factor (TNF)- α and/or through a direct cytolytic mechanism (Fig. 1). The mechanism of cytolytic killing can be mediated by the release of granule contents such as perforin and granzyme from CD8⁺ T cells. In addition, CD8⁺ T cells can kill cells by a process of Fas-mediated lysis.

The role of T_H1 and CD8⁺ T cells in protective immunity

In many experimental rodent models of intracellular infection, which include malaria and *Mycobacterium tuberculosis*, mice made deficient in genes for IFN- γ or IL-12 or functionally depleted of CD8⁺ T cells have all been shown to have an increased susceptibility to infection, with accelerated mortality. Moreover, humans with specific mutations in their receptors for IFN- γ or IL-12 have been shown to have enhanced susceptibility to mycobacterial disease. Together, these results provide strong evidence that for tuberculosis, malaria and other parasitic infections including *Leishmania major*, T_H1 and/or CD8⁺ T-cell responses are the effector mechanisms required for protective immunity.

Finally, there is evidence that CD8⁺ T cells also have a crucial role in mediating protection against HIV infection in both primates and humans. Thus, the development of a successful vaccine for these diseases will be facilitated by a thorough understanding of how cellular immune responses are generated and maintained *in vivo*.

Cellular memory

The major hallmarks of the immune response are discrimination between self and non-self, and immune memory. Whether it relates to cellular or humoral immunity, immune memory is the conceptual basis for how vaccines work. With regard to cellular memory, CD4⁺ and CD8⁺ memory T cells differ from naive T cells in the following manner. First, memory cells express specific cell-surface markers that can denote prior activation. Second, memory cells are qualitatively different from naive T cells and seem to have less stringent requirements for activation. In this regard, memory cells respond to a previously encountered antigen or pathogen (secondary response) more vigorously and faster^{1,2} than naive T cells. Furthermore, memory cells can be further divided into 'resting' and 'effector' cells. Effector memory cells are usually large, whereas resting memory cells are small. Effector CD4⁺ T cells (T_H1 cells) readily secrete cytokines, whereas effector CD8⁺ T cells have direct cytolytic activity *ex vivo*. By contrast, resting memory CD4⁺ or CD8⁺ T cells need to be stimulated with antigen before producing cytokines or acquiring cytolytic activity, respectively. Finally, a recent study³ suggests that differential expression of the chemokine receptor CCR7 segregates human memory T cells into two functionally distinct subsets. CCR7⁺ cells have been referred to as 'central' memory cells and express lymph-node homing receptors but lack immediate effector function. CCR7⁻ cells display immediate effector function and have been referred to as 'effector' memory cells. Thus, the differential expression of CCR7 on T cells permits the existence of two separate pools of memory cells that are dynamic and can serve to mediate effector function in response to infection at different sites. It should also be noted that, after activation, CCR7⁺ cells differentiate into CCR7⁻ cells and acquire effector function. These results support a linear differentiation model in which naive T cells first develop into central memory (CCR7⁺) and then proceed to effector memory cells (CCR7⁻)³. Two crucial factors that are most likely to regulate the generation, maintenance and effector aspects of T-cell memory responses and the transition from CCR7⁺ to CCR7⁻ cells are antigen and cytokines.

Requirement for antigen in maintaining cellular memory

In recent years there has been tremendous interest in whether antigen is required to sustain memory T cells. This issue has been controversial in that some reports show that antigen is required to maintain memory cells⁴⁻⁶, whereas others find that memory CD8⁺ T cells can be maintained in an antigen-independent manner⁷⁻⁹. It is currently our view, on the basis of all the evidence, that antigen is not essential to sustain memory cells *in vivo*. Moreover, in terms of cytokine production (namely IFN- γ) for both CD4⁺ and CD8⁺ T cells^{8,9}, rapid responsiveness can also be maintained from long-lived memory T cells in the absence of antigen. But as it relates to immune protection in diseases in which CD8⁺ T cells are important (such as malaria and HIV) there is strong evidence that antigen might be required to mediate a protective response. This could relate to the factors discussed below and is supported by the following results. First, in the rodent model of malaria infection after inoculation, parasites quickly enter the liver and infect hepatocytes. There is a marked expansion in parasite growth over the next several days, and then parasites exit the liver and enter the circulation, where they infect red blood cells. Because parasites infect hepatocytes, which express MHC class I, they are targets for CD8⁺ T-cell cytolytic activity. In this regard in a rodent model of malaria infection, after vaccination with DNA encoding specific malarial antigens, protection was induced in a manner dependent on CD8⁺ T cells and IFN- γ (ref. 10). These and other results clearly establish the importance of CD8⁺ T cells in protection against malaria in this rodent model. In addition, vaccination with

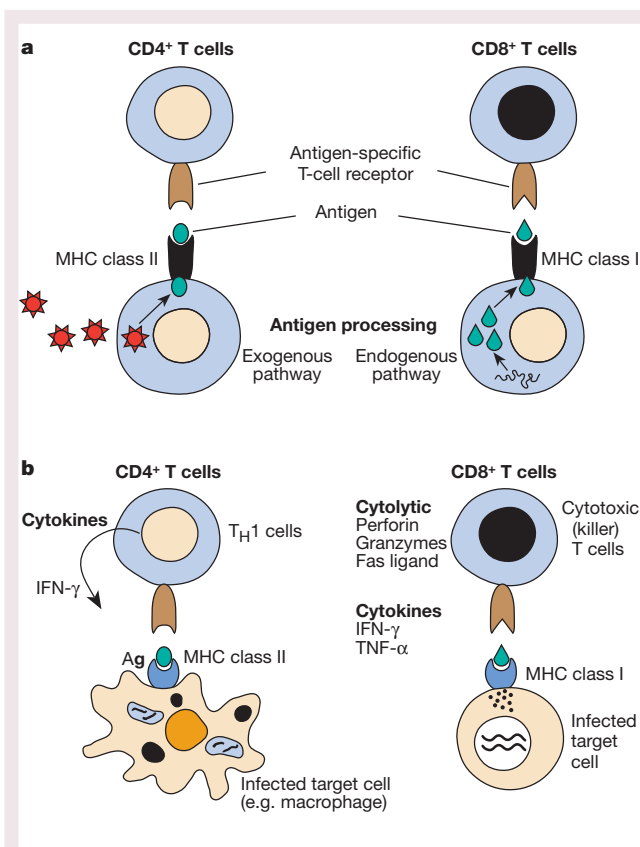


Figure 1 Mechanism of T-cell activation and effector function. **a**, Mechanism of antigen processing and recognition by T cells. **b**, Effector function of T_H1 and CD8⁺ T cells. Ag, antigen.

live irradiated sporozoites confers protection in rodents, primates and humans against a challenge with live unirradiated sporozoites¹¹. Protective immunity after vaccination with irradiated sporozoites in rats was abrogated by treatment with a drug that eliminated the pre-existing parasites in the liver before re-challenge¹². Thus, although the immune mechanism by which irradiated sporozoites are mediating their protective effect is not clear at present, these data suggest that persistent antigen might be required to maintain protection. Moreover, one could speculate that in sporozoite-vaccinated mice, persistent antigen (that is, sporozoites) was required to sustain CD8⁺ T-cell responses that were mediating protection. In infection with HIV, there is also strong correlative evidence that enhanced CD8⁺ T cell cytolytic activity is associated with a reduced viral load in infected primates^{13,14} and in patients with HIV¹⁵. Moreover, by monitoring the frequency of HIV-specific CD8⁺ T cells using MHC class I-specific tetramers, it was shown that treatment of patients with highly active antiretroviral therapy (HAART) caused a decrease in the number of antigen-specific CD8⁺ cells^{15,16}. This potential ability of antigen (viral load) to sustain CD8⁺ cytolytic activity is the basis for current clinical trials in which HIV-infected patients on HAART have structured interruptions of their treatment to expose the immune system briefly to antigen and thus sustain the cytolytic response^{16,17}. It is hoped that this strategy will generate a persistent cellular immune response that can serve to limit viral replication even after therapy ceases.

With regard to CD4⁺ T cell memory (T_H1 memory), it has been demonstrated that after the adoptive transfer of short-term T_H1 lines that were generated *in vitro*, cells are maintained *in vivo* for several months in the absence of antigen or even MHC class II molecules¹⁸. Thus, from these results, it is clear that antigen is not required for sustaining a pool of 'resting' memory T_H1 cells that were generated *in*

vitro. Moreover, in a different adoptive transfer model, it was also demonstrated that CD4 memory T cells could be sustained in the absence of antigen¹⁹. In this study, it was of interest that not all memory T cells expressed IFN- γ when assessed *ex vivo*. These results suggested that there could be a heterogeneous population of memory CD4⁺ T cells. One type of memory cell could rapidly produce IFN- γ after stimulation, whereas the other type of memory cell required additional differentiation for effector function. These results might be analogous to the delineation of effector/memory cells (CCR7⁻) and central/memory cells (CCR7⁺), respectively, as discussed above. What has not been well defined is whether persistent antigen is required for the maintenance of protective levels of memory/effector T_H1 cells that are generated *in vivo* after immunization. Some insight into the importance of antigen in maintaining immunity can be gained from observations in the course of human clinical disease. As noted above, T_H1 cells have a clear role in mediating protection in rodent models of *M. tuberculosis* and cutaneous infection with *Leishmania*. In most people with tuberculosis or cutaneous infection with *Leishmania*, once infected individuals have resolved their initial infection they remain resistant to reinfection. Moreover, reactivation of tuberculosis can occur long after the resolution of primary infection in immunocompromised subjects, suggesting the existence of persistent live mycobacteria. In addition, persistent antigen has been documented in cutaneous infection with *Leishmania* many years after primary infection²⁰. Taken together, these results suggest that persistent antigen might be important in sustaining immunity in diseases that require T_H1 cells and/or CD8⁺ cytolytic cells. Finally, although persistent antigen might be important for long-term immunity that is induced by primary infection or by vaccination for the diseases discussed here, there is also substantial evidence indicating that too much antigen can markedly impair T-cell immune responses. This could occur quantitatively by deletion of activated cells through antigen-induced cellular death (apoptosis) and/or qualitatively by inducing unresponsiveness of the T cell (clonal anergy). Thus, although the persistence of antigen might be important in conferring immune protection after vaccination, the amount of antigen present might also be a crucial factor in optimizing long-term T-cell immunity.

Requirement for cytokines in maintaining cellular memory

Another factor besides antigen that is crucial for generating and sustaining memory T cells is the role of cytokines. For CD8⁺ T cells, there is evidence that IFN- α acting through IL-15 might be important in sustaining memory CD8⁺ T-cell responses²¹. As noted above, CD4⁺ T cells require IL-12 for differentiation into functional T_H1 cells. Moreover, there is now evidence that IL-12 might also be required to sustain memory/effector T_H1 cells sufficient to mediate a biological outcome. This was shown in a mouse vaccine model of infection with *L. major* in which it was demonstrated that continuous IL-12 was required to sustain T_H1 cells sufficient to control infection after challenge²². Finally, T_H1 responses have also been shown to mediate protection against *M. tuberculosis*. Currently, the live attenuated Bacille Calmette–Guérin (BCG) strain, which is an inducer of IL-12 and T_H1 responses, does provide some protection against systemic disease early in childhood. By contrast, the efficacy of BCG vaccination against pulmonary disease occurring later in life is quite variable. Although there are many reasons why BCG might not be uniformly effective against pulmonary disease, one potential explanation is that, over time, there is a sterilization of BCG. This would limit the ability both to provide a source of antigen and to induce cytokines such as IL-12.

Thus, for both T_H1 and CD8⁺ T-cell memory responses, the persistence of antigen might not be required for the maintenance of 'resting' memory CD4⁺ or CD8⁺ T cells but might be crucial for sustaining a pool of 'effectors' sufficient to provide protection for those diseases that require cellular immunity. Moreover, cytokines (such as IL-12) might also be required for the maintenance of a pool of memory/effector T_H1 cells sufficient for protection. The exact

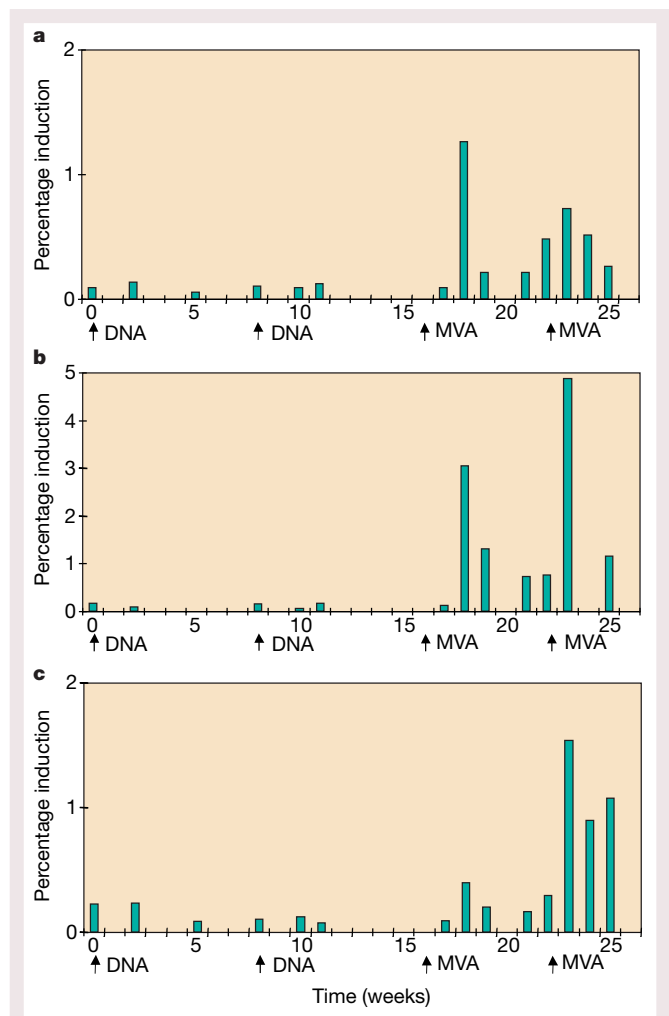


Figure 2 Induction of CD8⁺ T cells in three macaque monkeys (a–c) with the use of a heterologous prime–boost immunization regime (modified from ref. 46). Macaques were immunized with a plasmid DNA vaccine encoding a SIV CD8 T-cell epitope restricted by the MHC molecule Mamu-A*01 at weeks 0 and 8, and then immunized with a recombinant modified vaccinia Ankara (MVA) vaccine encoding the same epitope at weeks 16 and 22. The time course (in weeks) of induction of CD8 T cells specific for that epitope was measured by using an MHC–peptide tetramer; induction is expressed as the percentage of total peripheral blood CD8⁺ T cells. Prime–boosted animals showed higher responses than macaques immunized twice with MVA alone⁴⁶.

nature of these requirements depends on additional factors highlighted below.

Immune protection — thresholds, time and place

Thresholds

Most current vaccines require multiple immunizations for the induction of sufficient antibody to mediate protective immunity. In animal models of diseases requiring cellular immunity, there also seems to be a certain threshold of effector T_H1 or CD8⁺ T cells that is required for protection. For example, in rodent malaria a threshold level of about 400 IFN- γ -secreting peptide-specific CD8⁺ T cells per million splenocytes is required to protect against sporozoite challenge²³. In addition to the magnitude or threshold of the cellular response, it is also important to understand the ontogeny of the immune response after immunization or infection. In this regard, the induction of antigen-specific CD4⁺ and CD8⁺ T cells *in vivo* proceeds through three phases. First, after immunization or exposure to natural infection, there is an initial expansion phase in which there is a substantial increase in the frequency of antigen-specific cells^{24,25}.

Table 1 Comparative analysis of various vaccine formulations

		DNA vaccine	Live attenuated	Killed protein subunit	Live vector (e.g. poxvirus)
Immune response					
Humoral	B cells	++	+++	+++	++
Cellular	CD4 ⁺	+++ T _H 1*	± T _H 1	± T _H 1	+
	CD8 ⁺	++	+++	—	+++
Antigen presentation	MHC class	I and II	I and II	II	I and II
Memory					
Humoral		+++	+++	+++	++
Cellular		++	+++	±	++
Manufacturing					
Ease of development and production		++++	+	++	+++
Cost		+++	+	+	++++
Transport/storage		+++	+	+++	+
Safety		+++†	++‡	++++	++

*T_H2 responses can be induced by gene-gun immunization in mice.

†Data available only from phase I trials.

‡Live/attenuated vaccines may be precluded for use in immunocompromised patients and certain infections such as HIV.

This expansion phase, which usually occurs during the first week, is followed by the death phase, in which more than 90% of the activated cells undergo activation-induced apoptosis²⁶. For CD8⁺ T cells, the initial frequency of antigen-specific cells seems to be proportional to the antigenic load. In general, live infection induces a greater increase in the frequency of antigen-specific cells than immunizations with specific peptide, DNA vaccines or recombinant non-replicating viruses encoding the specific antigen²⁴. Moreover, it has been suggested that the persistence of antigen²⁷ and/or the amount of antigen²⁸ also affects qualitative aspects of CD8⁺ T-cell memory/effector responses. Once the initial expansion and death phases have occurred, the third phase of the response — the memory phase — can be stably maintained over a long period, even in the absence of persistent antigen^{1,7,29}. This memory phase of the response in the presence or absence of persistent antigen is likely to be heterogeneous, consisting of 'central' and 'effector' memory cells. The requirement for antigen in sustaining these specific memory cell types is not clear at present and is likely to have a large effect on whether or not immune protection is achieved, depending on the particular type of infection. Finally, the magnitude of the memory phase is generally determined by the size of the initial clonal burst induced by immunization³⁰. Thus, for protective memory CD8⁺ T-cell responses for diseases such as HIV and malaria that might require a high number of memory/effector cells, it is desirable to use a vaccine regimen that provides a sufficient antigenic load to maximize the initial burst size, so that a threshold of memory cells is maintained that will be protective.

Time and place

Currently used vaccines work by either preventing infection or preventing disease. For certain infections such as HIV, which can remain latent for long periods before progressing to clinical infection, the prevention of any infection by a vaccine would be optimal. Prevention of primary infection is probably mediated by a humoral immune response. By contrast, CD8⁺ T cells exert their effector function by recognizing antigen that has been processed in the context of MHC class I, and act to kill infected cells and to prevent a further spread of the infection. Thus, although CD8⁺ T cells might not themselves be able to prevent infection, they still could have a beneficial role in limiting or preventing disease, depending on how efficient they are in killing all of the infected cells. Furthermore, the effectiveness of the cellular immune response depends on where exposure to the pathogen occurs, and on whether the immune response exists within that site or needs to migrate from an immune organ. This crucial factor of time and place is determined by the specific infection. Thus, for rapidly replicating viruses such as influenza, it has been shown that the ability to generate long-term memory CD8⁺ T cells after primary infection is not sufficient to prevent disease in the lungs

after re-challenge³¹. In this study, even though the frequency and kinetics of the secondary CD8⁺ T-cell responses in the lungs were enhanced, it still took several days for cells to localize to the lungs from the mediastinal lymph nodes. With regard to HIV infection, it has recently been shown that after the intravaginal inoculation of rhesus monkeys with simian immunodeficiency virus (SIV), it was possible to find CD3⁺ T cells in the cervix that contained viral RNA by 3 days after infection³². These results imply a relatively narrow time window during which effector CD8⁺ T cells need either to be present at the mucosal site or to migrate quickly from a systemic site. At present, generating a high frequency of effector CD8⁺ cytolytic responses at mucosal sites has been difficult and represents a daunting challenge for vaccines. Nevertheless, the ability to generate a pool of 'systemic' memory CD8⁺ T cells that can be activated after exposure to HIV might not prevent infection or kill all infected cells, but could still limit the initial viral set point after acute infection, providing some measure of protection³³. In contrast with certain viral infections in which infection, latency and clinical disease can be established relatively quickly, time might be less of a limiting factor for protection against certain parasitic or mycobacterial infections. For malaria, the pre-erythrocytic liver stage of infection lasts for about 7 days in humans, providing a potentially long window for the action of effector T cells. It is important to note that complete killing of all parasites must occur within the liver to prevent release into the circulation and the infection of red blood cells. Finally, for pulmonary tuberculosis or cutaneous leishmanial infection, the time between initial exposure and clinical disease can be several weeks or longer. In these infections, it might be possible for a low frequency of 'resting' memory T cells that are generated after vaccination to develop into effector T cells that can control and prevent clinical disease. The challenges that lie ahead for vaccine development will be not only to optimize the qualitative and quantitative cellular immune response but also to induce responses directly at sites of infection.

Vaccines inducing cellular immunity to intracellular pathogens

All currently licensed vaccines (killed or inactivated, whole-cell, recombinant protein, or live attenuated) generate humoral immune responses. A potential shortcoming of non-live vaccines is their relative inefficiency in generating cellular immune responses (T_H1 or CD8⁺ T-cell responses), thus limiting their application for those diseases requiring cellular immunity (Table 1). Moreover, although certain adjuvants (used to enhance the potency of the vaccine) given with non-live vaccines have been shown to induce T_H1 responses, it has been difficult to find any adjuvant that can facilitate CD8⁺ T-cell induction in humans. In contrast, live attenuated vaccines permit the efficient MHC class I presentation of antigen that can stimulate CD8⁺ T-cell responses. The ability of live attenuated vaccines to induce

T_H1 -specific responses in humans, however, is less well characterized and can depend on the specific pathogen. Live vaccines also have the advantage of antigen persistence. From a practical standpoint, and also from one of safety, attenuated whole-organism vaccines raise several issues that can preclude their widespread application for certain diseases (namely HIV and malaria). Thus, the demonstration that DNA vaccines could induce both humoral and broad cellular responses (cytolytic and T_H1 responses)³⁴ in small animals provided hope that they could become the future vaccine of choice for infections requiring cellular immunity in humans³⁵. This optimism must be tempered by the fact that the induction of humoral and cellular immune responses after vaccination of humans with DNA has been disappointing in comparison with the rodent models^{36–38}. This has prompted investigators to try to optimize the immunogenicity of plasmid DNA vaccines themselves and/or to use DNA vaccines in combination with other vaccines to enhance immunity.

In terms of optimizing DNA vaccines in humans, several approaches are being taken to improve immunogenicity. In this regard, one approach has focused on specific nucleotide sequences (CpG motifs) within the plasmid. In both mice and humans, CpG oligodeoxynucleotides have been shown to stimulate multiple types of immune cell, leading to enhanced T_H1 and cytolytic $CD8^+$ T-cell responses^{39–42}. Moreover, recent work has shown that the immunostimulatory activity of CpG oligodeoxynucleotides can be blocked by certain non-CpG motifs⁴³. Thus, there is currently an effort to determine whether the addition of specific stimulatory CpG motifs, and the deletion of inhibitory ones, within the plasmid backbone can improve their immunogenicity. With regard to combining DNA with other types of vaccine, there have been several studies evaluating prime-boost strategies with vaccines based on live virus on or protein. Live virus vectors themselves typically generate stronger cellular immune responses than do DNA vaccines in small animals, yet clinical trials of poxvirus and adenovirus vectors have generally yielded fairly small cellular immune responses. The enhanced immunogenicity of a heterologous prime-boost immunization strategy, by using DNA priming and boosting with a recombinant poxvirus vector that encodes the same foreign antigens, has been demonstrated in several infectious diseases^{23,44}. In mice and in primates (Fig. 2), substantial enhancement of $CD8^+$ T-cell responses was observed, and the immunization order was crucial for maximal immunogenicity. Several poxviruses, such as modified vaccinia virus Ankara (MVA) and fowlpox, and also replication-defective adenoviruses, have this capacity to boost a primed $CD8^+$ T-cell response substantially^{45,46}. Clinical trials of this approach for malaria with MVA as a boosting agent are in progress. Finally, although the current approach of using a DNA prime and a MVA boost is clearly important in optimizing the magnitude of the $CD8^+$ response, it is not clear how long the enhanced responses will be maintained after the MVA boost. But this raises the issue of whether continued boosting will be required for sustaining a sufficient number of cells to confer protection.

Another method for optimizing the qualitative and quantitative immune response after vaccination is to use an adjuvant. Currently, the only widely used adjuvant in humans is alum, which improves antibody production but has little effect on T_H1 or $CD8^+$ T-cell responses. A novel adjuvant consisting of QS21, monophosphoryl lipid A (MPL) and an oil-in-water emulsion, SBAS2, has been shown to generate both strong antibody responses and T_H1 -type responses in combination with a protein antigen in malaria. This 'RTS,S' vaccine has shown impressive but relatively short-lived protective efficacy in sporozoite challenge studies and is currently undergoing field trials⁴⁷.

Perspective

Various new vaccination strategies for inducing strong cellular immune responses are now reaching clinical trials. Several of these are being evaluated for diseases such as malaria or HIV or in a

therapeutic setting where efficacy can be assessed in small studies. In addition, as noted above, other immunization approaches will probably be required for inducing effective mucosal immunity and for generating different types of T-cell response such as $CD1$ - and other non-classically restricted T cells. Although not all of these new approaches will generate sufficiently durable immunity for useful immunoprophylaxis, even the short-term potent induction of T effector cells might be of value for therapeutic vaccination. Furthermore, even if vaccines fail to induce long-term cellular immune responses, if exposure occurs within a short time of vaccination — a period during which the immune response is robust — protection could be achieved, and then life-long immunity would be possible as a result of the boosting effect of the live infection. The progress in developing vaccines for cellular immunity will also be enhanced by recent advances in the monitoring of T-cell responses by using sensitive enzyme-linked immunospot techniques and new tetrameric reagents. These tools provide excellent opportunities for the identification of immunological correlates of protection after vaccination. As a whole, our increased understanding of the immune correlates of protection as well as of the requirements for generating and sustaining cellular immune responses provides a framework for successful vaccine development against those diseases that require cellular immunity. □

- Ahmed, R. & Gray, D. Immunological memory and protective immunity: understanding their relation. *Science* **272**, 54–60 (1996).
- Dutton, R. W., Bradley, L. M. & Swain, S. L. T cell memory. *Annu. Rev. Immunol.* **16**, 201–223 (1998).
- Sallusto, F., Lenig, D., Forster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).
- Gray, D. & Matzinger, P. T cell memory is short-lived in the absence of antigen. *J. Exp. Med.* **174**, 969–974 (1991).
- Bachmann, M. F., Kundig, T. M., Hengartner, H. & Zinkernagel, R. M. Protection against immunopathological consequences of a viral infection by activated but not resting cytotoxic T cells: T cell memory without 'memory T cells'? *Proc. Natl Acad. Sci. USA* **94**, 640–645 (1997).
- Kundig, T. M. et al. On the role of antigen in maintaining cytotoxic T-cell memory. *Proc. Natl Acad. Sci. USA* **93**, 9716–9723 (1996).
- Lau, L. L., Jamieson, B. D., Somasundaram, T. & Ahmed, R. Cytotoxic T-cell memory without antigen. *Nature* **369**, 648–652 (1994).
- Tanchot, C., Lemonnier, F. A., Perarnau, B., Breit, A. A. & Rocha, B. Differential requirements for survival and proliferation of $CD8$ naive or memory T cells. *Science* **276**, 2057–2062 (1997).
- Murali-Krishna, K. et al. Persistence of memory $CD8$ T cells in MHC class I-deficient mice. *Science* **286**, 1377–1381 (1999).
- Doolan, D. L. et al. Circumventing genetic restriction of protection against malaria with multigene DNA immunization: $CD8^+$ cell-, interferon γ -, and nitric oxide-dependent immunity. *J. Exp. Med.* **183**, 1739–1746 (1996).
- Guebre-Xabier, M., Schwenk, R. & Krzych, U. Memory phenotype $CD8^+$ T cells persist in livers of mice protected against malaria by immunization with attenuated *Plasmodium berghei* sporozoites. *Eur. J. Immunol.* **29**, 3978–3986 (1999).
- Scheller, L. F. & Azad, A. F. Maintenance of protective immunity against malaria by persistent hepatic parasites derived from irradiated sporozoites. *Proc. Natl Acad. Sci. USA* **92**, 4066–4068 (1995).
- Matano, T. et al. Administration of an anti- $CD8$ monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. *J. Virol.* **72**, 164–169 (1998).
- Schmitz, J. E. et al. Control of viremia in simian immunodeficiency virus infection by $CD8^+$ lymphocytes. *Science* **283**, 857–860 (1999).
- Ogg, G. S. et al. Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103–2106 (1998).
- Kalams, S. A. et al. Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. *J. Virol.* **73**, 6721–6728 (1999).
- Ortiz, G. M. et al. HIV-1-specific immune responses in subjects who temporarily contain virus replication after discontinuation of highly active antiretroviral therapy. *J. Clin. Invest.* **104**, R13–R18 (1999).
- Swain, S. L., Hu, H. & Huston, G. Class II-independent generation of $CD4$ memory T cells from effectors. *Science* **286**, 1381–1383 (1999).
- Garcia, S., DiSanto, J. & Stockinger, B. Following the development of a $CD4$ T cell response *in vivo*: from activation to memory formation. *Immunity* **11**, 163–171 (1999).
- Schuback, A. et al. Detection of *Leishmania* DNA by polymerase chain reaction in scars of treated human patients. *J. Infect. Dis.* **178**, 911–914 (1998).
- Zhang, X., Sun, S., Hwang, L., Tough, D. F. & Sprent, J. Potent and selective stimulation of memory-phenotype $CD8^+$ T cells *in vivo* by IL-15. *Immunity* **8**, 591–599 (1998).
- Gurunathan, S., Prussin, C., Sacks, D. L. & Seder, R. A. Vaccine requirements for sustained cellular immunity to an intracellular parasitic infection. *Nature Med.* **4**, 1409–1415 (1998).
- Schneider, J. et al. Enhanced immunogenicity for $CD8^+$ T cell induction and complete protective efficacy of malaria DNA vaccination by boosting with modified vaccinia virus Ankara. *Nature Med.* **4**, 397–402 (1998).
- Oehen, S., Waldner, H., Kundig, T. M., Hengartner, H. & Zinkernagel, R. M. Antivirally protective cytotoxic T cell memory to lymphocytic choriomeningitis virus is governed by persisting antigen. *J. Exp. Med.* **176**, 1273–1281 (1992).

25. McHeyzer-Williams, M. G. & Davis, M. M. Antigen-specific development of primary and memory T cells *in vivo*. *Science* **268**, 106–111 (1995).
26. Razvi, E. S. & Welsh, R. M. Apoptosis in viral infections. *Adv. Virus Res.* **45**, 1–60 (1995).
27. Ochsenbein, A. F. *et al.* A comparison of T cell memory against the same antigen induced by virus versus intracellular bacteria. *Proc. Natl Acad. Sci. USA* **96**, 9293–9298 (1999).
28. Opferman, J. T., Ober, B. T. & Ashton-Rickardt, P. G. Linear differentiation of cytotoxic effectors into memory T lymphocytes. *Science* **283**, 1745–1748 (1999).
29. Mullbacher, A. The long-term maintenance of cytotoxic T cell memory does not require persistence of antigen. *J. Exp. Med.* **179**, 317–321 (1994).
30. Hou, S., Hyland, L., Ryan, K. W., Portner, A. & Doherty, P. C. Virus-specific CD8⁺ T-cell memory determined by clonal burst size. *Nature* **369**, 652–654 (1994).
31. Flynn, K. J. *et al.* Virus-specific CD8⁺ T cells in primary and secondary influenza pneumonia. *Immunity* **8**, 683–691 (1998).
32. Zhang, Z. *et al.* Sexual transmission and propagation of SIV and HIV in resting and activated CD4⁺ T cells. *Science* **286**, 1353–1357 (1999).
33. Lifson, J. D. *et al.* The extent of early viral replication is a critical determinant of the natural history of simian immunodeficiency virus infection. *J. Virol.* **71**, 9508–9514 (1997).
34. Ulmer, J. B. *et al.* Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* **259**, 1745–1749 (1993).
35. Seder, R. A. & Gurunathan, S. DNA vaccines — designer vaccines for the 21st century. *N. Engl. J. Med.* **341**, 277–278 (1999).
36. Wang, R. *et al.* Induction of antigen-specific cytotoxic T lymphocytes in humans by a malaria DNA vaccine. *Science* **282**, 476–480 (1998).
37. Le, T. P. *et al.* Safety, tolerability and humoral immune responses after intramuscular administration of a malaria DNA vaccine to healthy adult volunteers. *Vaccine* **18**, 1893–1901 (2000).
38. Calarota, S. *et al.* Cellular cytotoxic response induced by DNA vaccination in HIV-1-infected patients. *Lancet* **351**, 1320–1325 (1998).
39. Cho, H. J. *et al.* Immunostimulatory DNA-based vaccines induce cytotoxic lymphocyte activity by a T-helper cell-independent mechanism. *Nature Biotechnol.* **18**, 509–514 (2000).
40. Cowdery, J. S., Chace, J. H., Yi, A. K. & Krieg, A. M. Bacterial DNA induces NK cells to produce IFN- γ *in vivo* and increases the toxicity of lipopolysaccharides. *J. Immunol.* **156**, 4570–4575 (1996).
41. Klinman, D. M., Yi, A. K., Beaucage, S. L., Conover, J. & Krieg, A. M. CpG motifs present in bacterial DNA rapidly induce lymphocytes to secrete interleukin 6, interleukin 12, and interferon γ . *Proc. Natl Acad. Sci. USA* **93**, 2879–2883 (1996).
42. Sato, Y. *et al.* Immunostimulatory DNA sequences necessary for effective intradermal gene immunization. *Science* **273**, 352–354 (1996).
43. Krieg, A. M. *et al.* Sequence motifs in adenoviral DNA block immune activation by stimulatory CpG motifs. *Proc. Natl Acad. Sci. USA* **95**, 12631–12636 (1998).
44. Schneider, J. *et al.* Induction of CD8⁺ T cells using heterologous prime-boost immunisation strategies. *Immunol. Rev.* **170**, 29–38 (1999).
45. Kent, S. J. *et al.* Enhanced T-cell immunogenicity and protective efficacy of a human immunodeficiency virus type 1 vaccine regimen consisting of consecutive priming with DNA and boosting with recombinant fowlpox virus. *J. Virol.* **72**, 10180–10188 (1998).
46. Hanke, T. *et al.* Effective induction of simian immunodeficiency virus-specific cytotoxic T lymphocytes in macaques by using a multiepitope gene and DNA prime-modified vaccinia virus Ankara boost vaccination regimen. *J. Virol.* **73**, 7524–7532 (1999).
47. Stoute, J. A. *et al.* A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **336**, 86–91 (1997).

Correspondence and requests for materials should be addressed to R.A.S. (e-mail: rseder@niaid.nih.gov).

Microbial genome sequencing

Claire M. Fraser, Jonathan A. Eisen & Steven L. Salzberg

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

Complete genome sequences of 30 microbial species have been determined during the past five years, and work in progress indicates that the complete sequences of more than 100 further microbial species will be available in the next two to four years. These results have revealed a tremendous amount of information on the physiology and evolution of microbial species, and should provide novel approaches to the diagnosis and treatment of infectious disease.

Microbes were the first organisms on Earth and preceded animals and plants by more than 3 billion years. They are the foundation of the biosphere, from both an evolutionary and an environmental perspective¹. It has been estimated that microbial species comprise about 60% of the Earth's biomass. The genetic, metabolic and physiological diversity of microbial species is far greater than that found in plants and animals. But the diversity of the microbial world is largely unknown, with less than one-half of 1% of the estimated 2–3 billion microbial species identified. Of those species that have been described, their biological diversity is extraordinary, having adapted to grow under extremes of temperature, pH, salt concentration and oxygen levels.

Perhaps no other area of research has been so energized by the application of genomic technology than the microbial field. It was only five years ago that The Institute for Genomic Research (TIGR) published the first complete genome sequence for a free-living organism, *Haemophilus influenzae*²; since that first report another 27 microbial genome sequences have been published, with at least 10–20 other projects at or near completion (for details see <http://www.tigr.org/tldb/mdb/mdb.html>). This progress represents, on average, one completed genome sequence every two months and all indications are that this pace will continue to accelerate. Included in the first completed microbial projects are many important human pathogens, the simplest known free-living organism, 'model' organisms, *Escherichia coli* and *Bacillus subtilis*, thermophilic bacterial species that might represent some of the deepest-branching members of the bacterial lineage, five representatives of the archaeal domain, and the first eukaryote, *Saccharomyces cerevisiae*. All of the organisms that have been studied by whole-genome analysis are species that can be grown either in the laboratory or in animal cells. It is important to remember that the vast majority of microbial species cannot be cultivated at all, and these organisms, which live in microbial communities, are essential to the overall ecology of the planet. Nevertheless, the study of 'laboratory-adapted' microbes has had a profound impact on our understanding of the biology and the evolutionary relationships between microbial species.

Methods for whole-genome analysis

The method that was successfully used to determine the complete genome sequence of *H. influenzae* is a shotgun sequencing strategy (Fig. 1). Before 1995, the largest genome sequenced with a random strategy was that of bacteriophage lambda with a genome size of 48,502 base pairs (bp), completed by Sanger *et al.* in 1982 (ref. 3). Despite

advances in DNA-sequencing technology, the sequencing of whole genomes had not progressed beyond lambda-sized clones (about 40 kbp) because of the lack of sufficient computational approaches that would enable the efficient assembly of a large number of independent random sequences into a single contig.

For the *H. influenzae* and subsequent projects, we have used a computational method that was developed to create assemblies from hundreds of thousands of complementary DNA sequences 300–500-bp long⁴. This approach has proved to be a cost-effective and efficient approach to sequencing megabase-sized segments of genomic DNA. This strategy does not require an ordered set of cosmids or other subclones, thus significantly reducing the overall cost per base pair of producing a finished sequence, while providing high redundancy for accuracy and minimizing the effort required to obtain the whole genome sequence. The availability of improved technologies for longer sequence lengths (more than 700 bp) reduces problems associated with repetitive elements in the final assembly.

Microbial gene finding and annotation

The identification of genes in prokaryotic genomes has advanced to the stage at which nearly all protein-coding regions can be identified with confidence. Computational gene finders using Markov modelling techniques now routinely find more than 99% of protein-coding regions⁵ and RNA genes⁶. Once the protein-coding genes have been located, the most challenging problem is to determine their function. Typically, about 40–60% of the genes in a newly sequenced bacterial genome display a detectable sequence similarity to protein sequences whose function is at least tentatively known. This sequence similarity is the primary basis for assigning function to new proteins, but the transfer of functional assignments is fraught with difficulties.

To illustrate this problem, Table 1 contains an example showing the best matches in the database for a 1,344-bp gene from *Mycobacterium tuberculosis* at the time that the genome was being sequenced. All six of the best matches are kinases, but the specific names differ. A conservative naming strategy might use a family name that includes all six matches. Another strategy might use curated protein families (if they exist) to assign names; for example, the FGGY family named in the fourth line of Table 1 comes from the Pfam database⁷, a set of 1,815 hidden Markov models based on multiple alignments. By a closer examination of the literature, one could determine which of these protein names were based on laboratory experiments and which on sequence similarity. In any case, the assignment of a function to this protein requires the expertise of a skilled biologist. The rapidly changing nature of genome databases

means that database searches must be repeated regularly to keep annotation accurate and up to date.

One possible solution to the annotation problem is to bring more of the resources of the scientific community to bear on each genome. No single centre can annotate all the functions of a living organism; experts from many different areas of biology should be encouraged to contribute to the annotation process. One possible model would be for geographically separated experts to deposit annotation to a central repository, which might also take on a curatorial or editorial role. An alternative model is one in which annotation resides in many different locations (as it does today), but in which new electronic links are created that allow scientists to locate rapidly all the information about a gene, genome or function. This latter model scales more easily and avoids the problem of overdependence on a single source.

What have we learned from genome analysis?

Comparison of the results from 24 completed prokaryotic genome sequences, containing more than 50 Mbp of DNA sequence and 54,000 predicted open reading frames (ORFs), has revealed that gene density in the microbes is consistent across many species, with about one gene per kilobase (Table 2). Almost half of the ORFs in each species are of unknown biological function. When the function of this large subset of genes begins to be explained, it is likely that entirely novel biochemical pathways will be identified that might be relevant to medicine and biotechnology. Perhaps even more unexpected is the observation that about a quarter of the ORFs in each species studied so far are unique, with no significant sequence similarity to any other available protein sequence. Although this might at present be an artefact of the small number of microbial species studied by whole-genome analysis, it nevertheless supports the idea that there is tremendous biological diversity between microorganisms. Taken together, these data indicate that much of microbial biology has yet to be understood and suggest that the idea of a 'model' organism in the microbial world might not be appropriate, given the vast differences between even related species.

Our molecular picture of evolution for the past 20 years has been dominated by the small-subunit ribosomal RNA phylogenetic tree

Table 1 Results of a BLAST search of a newly sequenced *M. tuberculosis* gene against a comprehensive protein database

Gene ID	Similarity (%)	Length (bp)	Gene name	E-value*
GP:2905647	44.8	1,191	D-Arabinitol kinase (<i>Klebsiella pneumoniae</i>)	6.2e-15
EGAD:22614	46.2	1,191	Gluconokinase (<i>Bacillus subtilis</i>)	1.4e-13
EGAD:20418	43.0	1,302	Xylulose kinase (<i>Lactobacillus pentosus</i>)	4.8e-13
EGAD:105114	43.4	1,320	Carbohydrate kinase, FGGY family (<i>Archaeoglobus fulgidus</i>)	4.7e-12
GP:2895855	42.7	1,263	Xylulokinase (<i>Lactobacillus brevis</i>)	1.0e-07
EGAD:10899	45.4	1,296	Xylulose kinase (<i>Escherichia coli</i>)	2.1e-06

*E-value is a statistical measure of the significance of a BLAST search result.

that proposes three non-overlapping groups of living organisms: the bacteria, the archaea and the eukaryotes⁸. Although the archaea possess bacterial cell structures, it has been suggested that they share a common ancestor exclusive of bacteria.

Analysis of complete genome sequences is beginning to provide great insight into many questions about the evolution of microbes. One such area has encompassed the occurrence of genetic exchanges between different evolutionary lineages, a phenomenon known as horizontal, or lateral, gene transfer. The occurrence of horizontal gene transfer, such as that involving genes from organellar genomes to the nucleus, or of antibiotic resistance genes between bacterial species, has been well established for many years (see, for example, ref. 9). This phenomenon causes problems for studying the evolution of species because it means that some species are chimaeric, with different histories for different genes. Before the availability of complete genome sequences, studies of horizontal gene transfer had been limited because of the incompleteness of the data sets being analysed. Analyses of complete genome sequences have led to many recent suggestions that the extent of horizontal gene exchange is much greater than was previously realized¹⁰⁻¹². For example, an

Table 2 Genome features from 24 microbial genome sequencing projects

Organism	Genome size (Mbp)	No. of ORFs (% coding)	Unknown function	Unique ORFs
<i>Aeropyrum pernix</i> K1	1.67	1,885 (89%)		
<i>A. aeolicus</i> VF5	1.50	1,749 (93%)	663 (44%)	407 (27%)
<i>A. fulgidus</i>	2.18	2,437 (92%)	1,315 (54%)	641 (26%)
<i>B. subtilis</i>	4.20	4,779 (87%)	1,722 (42%)	1,053 (26%)
<i>B. burgdorferi</i>	1.44	1,738 (88%)	1,132 (65%)	652 (39%)
<i>Chlamydia pneumoniae</i> AR39	1.23	1,134 (90%)	543 (48%)	262 (23%)
<i>Chlamydia trachomatis</i> MoPn	1.07	936 (91%)	353 (38%)	77 (8%)
<i>C. trachomatis</i> serovar D	1.04	928 (92%)	290 (32%)	255 (29%)
<i>Deinococcus radiodurans</i>	3.28	3,187 (91%)	1,715 (54%)	1,001 (31%)
<i>E. coli</i> K-12-MG1655	4.60	5,295 (88%)	1,632 (38%)	1,114 (26%)
<i>H. influenzae</i>	1.83	1,738 (88%)	592 (35%)	237 (14%)
<i>H. pylori</i> 26695	1.66	1,589 (91%)	744 (45%)	539 (33%)
<i>Methanobacterium thermotautotrophicum</i>	1.75	2,008 (90%)	1,010 (54%)	496 (27%)
<i>Methanococcus jannaschii</i>	1.66	1,783 (87%)	1,076 (62%)	525 (30%)
<i>M. tuberculosis</i> CSU#93	4.41	4,275 (92%)	1,521 (39%)	606 (15%)
<i>M. genitalium</i>	0.58	483 (91%)	173 (37%)	7 (2%)
<i>M. pneumoniae</i>	0.81	680 (89%)	248 (37%)	67 (10%)
<i>N. meningitidis</i> MC58	2.24	2,155 (83%)	856 (40%)	517 (24%)
<i>Pyrococcus horikoshii</i> OT3	1.74	1,994 (91%)	859 (42%)	453 (22%)
<i>Rickettsia prowazekii</i> Madrid E	1.11	878 (75%)	311 (37%)	209 (25%)
<i>Synechocystis</i> sp.	3.57	4,003 (87%)	2,384 (75%)	1,426 (45%)
<i>T. maritima</i> MSB8	1.86	1,879 (95%)	863 (46%)	373 (26%)
<i>T. pallidum</i>	1.14	1,039 (93%)	461 (44%)	280 (27%)
<i>Vibrio cholerae</i> El Tor N1696	4.03	3,890 (88%)	1,806 (46%)	934 (24%)
	50.60	52,462 (89%)	22,358 (43%)	12,161 (23%)

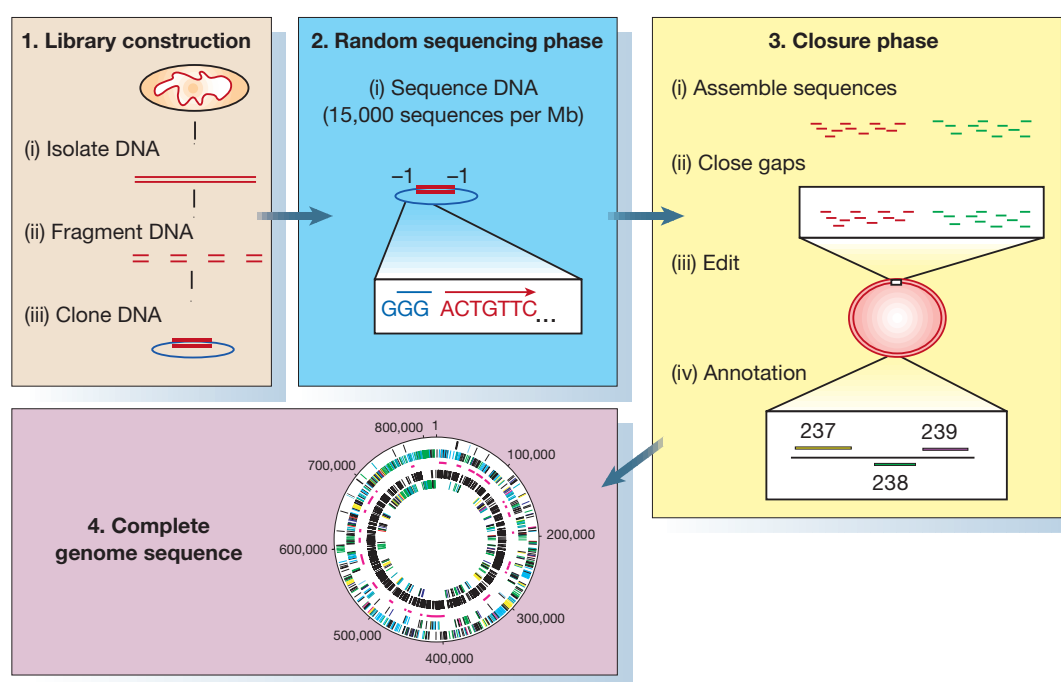


Figure 1 Diagram depicting the steps in a whole-genome shotgun sequencing project.

analysis of the genomes of two thermophilic bacterial species, *Aquifex aeolicus* and *Thermotoga maritima*, revealed that 20–25% of the genes in these species were more similar to genes from archaea than those from bacteria^{13,14}. This led to the suggestion of possible extensive gene exchanges between these species and archaeal lineages. But before one jumps to this conclusion it is important to consider the difficulties in inferring the occurrence of gene transfer. For example, the high percentage of genes with best matches to archaea in *A. aeolicus* and *T. maritima* could also be due to a high rate of evolution in the mesophilic bacteria (which would cause thermophilic and archaeal genes to have high levels of similarity despite their not having a common ancestry) or the loss of these genes from mesophilic bacteria¹⁵. For *T. maritima*, many lines of additional evidence support the assertion of gene transfer, including the observation that many of the archaeal-like genes occur in clusters in the genome, are in regions of unusual nucleotide composition, and branch in phylogenetic trees most closely to archaeal genes¹⁴. Most of the lines of evidence leading to assertions of horizontal gene transfer can have other causes. For example, unusual nucleotide composition can also arise from selection¹⁶, and differences in phylogenetic trees can be caused by convergence, inaccurate alignments¹⁷, long-branch attraction¹⁸ or sampling of different species¹⁹. It is therefore important to assess the evidence carefully and to find multiple types of evidence. This has yet to be done systematically, so we believe that it is too early to assign quantitative values to the extent of gene exchange between species.

Despite the apparent occurrence of extensive gene transfers in the history of microbes, it does seem that there might be a 'core' to each evolutionary lineage that retains some phylogenetic signal. The best evidence for this comes from the construction of 'whole genome trees' based on the presence and absence of particular homologues or orthologues in different complete genomes²⁰. It is important to note that gene-content trees are averages of patterns produced by phylogeny, gene duplication and loss, and horizontal transfer; they are therefore not real phylogenetic trees. Nevertheless, the fact that these trees are very similar to phylogenetic trees of genes such as ribosomal RNA and RecA suggests that although horizontal gene transfer might

be extensive, it is somehow constrained by phylogenetic relationships. Other evidence for a 'core' of particular lineages comes from the finding of a conserved core of euryarchaeal genomes^{21,22} and another finding that some types of gene might be more prone to gene transfer than others²³. It therefore seems likely that horizontal gene transfer has not completely obliterated the phylogenetic signal in microbial genomes. Careful studies in which the phylogenetic trees of some of these core genes are compared across all genomes need to be done to see whether or not the core has a consistent phylogeny. Initial studies suggest that it does, at least for the major microbial groups¹⁴.

Although our ability to resolve patterns of the relationships among microbes is still limited, analysis of the genomes of closely related species is revealing much about genome evolution^{24,25}. For example, a comparison of the genomes of four chlamydial species has revealed the occurrence of frequent tandem gene duplication and gene loss, as well as large chromosomal inversions²⁵. Comparisons of closely related species should also reveal much about mutation processes, codon usage and other features that evolve rapidly¹⁶.

Design of new antimicrobial agents and vaccines

One of the expected benefits of genome analysis of pathogenic bacteria is in the area of human health, particularly in the design of more rapid diagnostic reagents and the development of new vaccines and antimicrobial agents. These goals have become more urgent with the continuing spread of antibiotic resistance in important human pathogens. Moreover, results from the whole-genome analysis of human pathogens has suggested that there are mechanisms for generating antigenic variation in proteins expressed on the cell surface that are encoded within the genomes of these organisms. These mechanisms include the following: (1) slipped-strand mispairing within DNA sequence repeats found in 5'-intergenic regions and coding sequences as described for *H. influenzae*², *Helicobacter pylori*²⁶ and *M. tuberculosis*²⁷, (2) recombination between homologous genes encoding outer-surface proteins as described for *Mycoplasma genitalium*²⁸, *Mycoplasma pneumoniae*²⁹ and *Treponema pallidum*³⁰, and (3) clonal variability in surface-expressed proteins as described for *Plasmodium falciparum*³¹ and possibly *Borrelia burgdorferi*³².

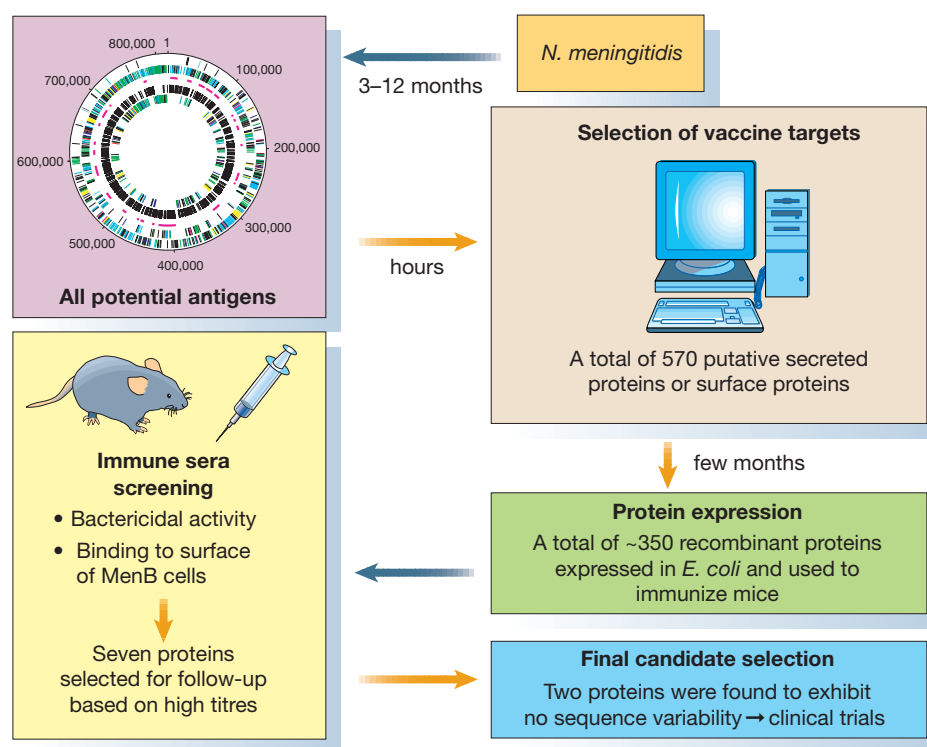


Figure 2 Diagram depicting how complete microbial genome sequence data can accelerate vaccine development.

Experimental evidence from studies of clinical isolates of some species has demonstrated phenotypic variation in the relevant cell-surface proteins³³, suggesting that, at least for human pathogens, the evolution of antigenic proteins probably occurs in real time, as cell populations divide. The ability of human pathogens to alter their antigenic potential and thereby evade the immune system has the potential to hinder vaccine development by conventional methods.

Progress during the past year has supported the idea that complete genome sequence information can be exploited in the design of new vaccines and antimicrobial compounds. As an example, the identification of new vaccine candidates against serogroup B *Neisseria meningitidis* (MenB) was reported by Pizza *et al.* using a genomics-based approach³⁴ (Fig. 2). With the use of the entire genome sequence of a virulent serogroup B strain³⁵, 570 putative cell-surface-expressed or secreted proteins were identified; the corresponding DNA sequences were cloned and expressed in *E. coli*. Of the putative targets, 61% were expressed successfully and used to immunize mice. Immune sera were screened for bactericidal activity and for the ability to bind to the surface of MenB cells. Seven representative proteins were selected for further study and were evaluated for their degree of sequence variability among multiple isolates and serogroups of *N. meningitidis*. Two highly conserved vaccine candidates emerged from this large-scale screening effort, which occurred in parallel with the completion of the genome sequence of *N. meningitidis*. These results provide the first definitive demonstration of the potential of genomic information to expand and accelerate the development of vaccines against pathogenic organisms.

Another example illustrates the potential of genomics to accelerate the development of novel antimicrobial agents. Jomaa *et al.*³⁶ identified two genes in *P. falciparum* from sequence data from the malaria genome consortium that encode key enzymes in the 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway that are required for the synthesis of isoprenoids such as cholesterol³⁷. The DOXP pathway functions in some bacteria, algae and higher plants to

produce isopentenyl diphosphate, a precursor of isoprenoids. In *P. falciparum*, the enzymes of the DOXP pathway are probably associated with a specialized organelle derived from algae called the apicoplast; they are expressed when the parasite is growing within red blood cells. Inhibitors of one of the key enzymes in the DOXP pathway, DOXP reductoisomerase, had previously been identified and had been shown to inhibit the bacterial enzyme and the growth of some bacterial species. Jomaa *et al.*³⁶ demonstrated that two inhibitors of DOXP reductoisomerase, fosidomycin and FR900098, were able to inhibit the growth of *P. falciparum* *in vitro* and cure mice infected with a related species of *Plasmodium*. Both of these compounds exhibit low toxicity and high stability and are relatively inexpensive to produce, suggesting that they might be the basis of a potentially important new class of anti-malarial drugs.

Conclusions

So far, studies in genomics have only scratched the surface of microbial diversity and have revealed how little is known about microbial species. In the next few years, more than 100 projects for sequencing microbial genomes should be completed, providing the scientific community with information on more than 300,000 predicted genes. A significant number of these genes will be novel and of unknown function. These novel genes represent exciting new opportunities for future research and potential sources of biological resources to be explored and exploited. The benefits of comparative genomics in understanding biochemical diversity, virulence and pathogenesis, and the evolution of species has been unequivocally demonstrated and the usefulness of comparative techniques will improve as more genomes become available. One of the major challenges is to develop techniques for assessing the function of novel genes on a large scale and integrating information on how genes and proteins interact at the cellular level to create and maintain a living organism. It is not unreasonable to expect that, by expanding our understanding of microbial biology and biodiversity, great strides can be made in the

diagnosis and treatment of infectious diseases and in the identification of useful functions in the microbial world that could be applied to agricultural and industrial processes.

1. Staley, J. J. *et al.* *The Microbial World. The Foundation of the Biosphere* (American Society for Microbiology, Washington DC, 1997).
2. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* **269**, 496–512 (1995).
3. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, S. Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.* **162**, 729 (1982).
4. Sutton, G. G., White, O., Adams, M. D. & Kerlavage, A. R. TIGR Assembler: a new tool for assembling large shotgun sequencing projects. *Genome Sci. Technol.* **1**, 9–19 (1995).
5. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with Glimmer. *Nucleic Acids Res.* **27**, 4636–4641 (1999).
6. Lowe, T. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
7. Bateman, A. *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **28**, 263–266 (2000).
8. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090 (1977).
9. Davies, J. Origins and evolution of antibiotic resistance. *Microbiology* **12**, 9–16 (1996).
10. Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417 (1998).
11. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2129 (1999).
12. Martin, W. Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104 (1999).
13. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* **14**, 442–444 (1998).
14. Nelson, K. E. *et al.* Genome sequencing of *Thermotoga maritima*: evidence for lateral gene transfer between bacteria and archaea. *Nature* **399**, 323–329 (1999).
15. Kyrpides, N. C. & Olsen, G. J. Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* **15**, 298–299 (1999).
16. Lafay, B. *et al.* Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Res.* **27**, 1642–1649 (1999).
17. Gatesy, J., Desalle, R. & Wheller, W. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylog. Evol.* **2**, 152–157 (1993).
18. Philippe, H. & Forterre, P. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49**, 509–523 (1999).
19. Eisen, J. A. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J. Mol. Evol.* **41**, 1105–1123 (1995).
20. Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature Genet.* **21**, 108–110 (1999).
21. Makarova, K. S. *et al.* Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* **9**, 608–628 (1999).
22. Graham, D. E., Overbeek, R., Olsen, G. J. & Woese, C. R. An archaeal genomic signature. *Proc. Natl Acad. Sci. USA* **97**, 3304–3308 (2000).
23. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 3801–3806 (1999).
24. Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen, *Helicobacter pylori*. *Nature* **397**, 176–180 (1999).
25. Read, T. D. *et al.* Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**, 1397–1406 (2000).
26. Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547 (1997).
27. Cole, S. T. *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544 (1998).
28. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
29. Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium, *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**, 4420–4429 (1996).
30. Fraser, C. M. *et al.* Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**, 375–388 (1998).
31. Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
32. Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586 (1997).
33. Peterson, S. N. *et al.* Characterization of repetitive DNA in the *Mycoplasma genitalium* genome: possible role in the generation of antigenic variation. *Proc. Natl Acad. Sci. USA* **92**, 11829–11833 (1995).
34. Pizzo, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole genome sequencing. *Science* **287**, 1816–1820 (2000).
35. Tettelin, H. *et al.* Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809–1815 (2000).
36. Jomaa, H. *et al.* Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576 (1999).
37. Ridley, R. G. Planting the seeds of new antimalarial drugs. *Science* **285**, 1502–1503 (1999).

BRISTOL-MYERS SQUIBB AND MICROBIAL DISEASE:

Basic Science, Clinical Development, Global Surveillance

Worldwide, more people die from infectious diseases than any other cause — more than 17 million in 1997, according to the World Health Organization.¹ The need to discover and develop new antimicrobial agents is critical to improving mankind's future health. Researchers taking up this challenge, however, do so under very difficult circumstances, such as growing antimicrobial resistance, the emergence of new pathogens, and the resurgence of age-old infectious diseases, including malaria and tuberculosis.

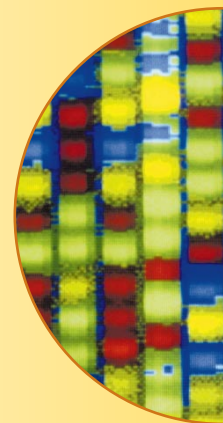
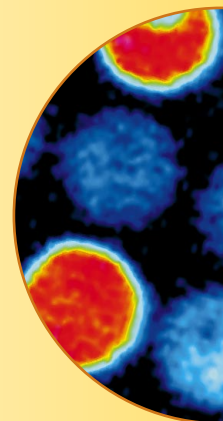
Researchers at Bristol-Myers Squibb Company have been a driving force behind the company's stature as a global leader in the area of infectious disease treatments. Our long-standing history in microbial disease research began in the 1940s with discoveries that helped make possible the mass production of penicillin. Today, our pharmaceutical research pipeline includes several potentially important compounds that span the infectious disease spectrum, including novel antibacterials, antifungals, and antivirals. And with applied genomics, pharmacogenomics, and clinical development programs in full swing, Bristol-Myers Squibb is taking research to a new level — to identify the right targets for the development of new and more effective antimicrobials, and turn the best targets into medicines that cure and treat disease.

Beyond discovering and developing antimicrobial therapies, Bristol-Myers Squibb wholly funds the SENTRY Antimicrobial Surveillance program. True to its name, SENTRY was established in 1997 to watch over and track the susceptibility patterns of bacterial and fungal pathogens throughout the world. SENTRY is the largest program of its kind, and provides the type of global and longitudinal surveillance of resistant pathogens that is critical to the timely revision of treatment strategies for microbial infections. This surveillance is also important for the proper prioritization of unmet medical needs that must be addressed by drug discovery efforts.

Despite the difficult challenges faced by infectious disease researchers today, we have great reason to be confident that many of the infectious diseases killing and disabling millions of people around the world — particularly in poorer regions — can be controlled if not eradicated in the foreseeable future. The articles in this *Nature Insight* capture important advances in our understanding of drug resistance, host defenses, antimicrobial peptides, and microbial comparative genomics, among other topics. At Bristol-Myers Squibb, we are already putting these and other learnings to good use as we pursue our mission to extend and enhance human life.

Peter S. Ringrose, Ph. D.

*Chief Scientific Officer, Bristol-Myers Squibb and
President, Pharmaceutical Research Institute*



Screen test

High-throughput screening gets put through its paces.

Isolute Array

International Sorbent Technology

www.ist-spe.com

A versatile 96-well solid-phase extraction system

Isolute Array SPE columns and 96-well extraction plates are designed for scientists who need a versatile SPE format compatible with both high and low sample throughput. Individual wells can either be extracted using conventional SPE vacuum manifolds or assembled by the end-user to form a customized extraction plate for method development and routine sample processing. Extraction plates can be processed using the VacMaster-96 sample processing station, which is compatible with both automated sample processors and manual operation. Fully populated plates or separate base plates and packs of Isolute SPE wells for self-assembly are available in 10, 25, 50 and 100 mg options. A range of sorbents, application development kits and associated generic methods are available to speed up the development of drug extraction procedures.

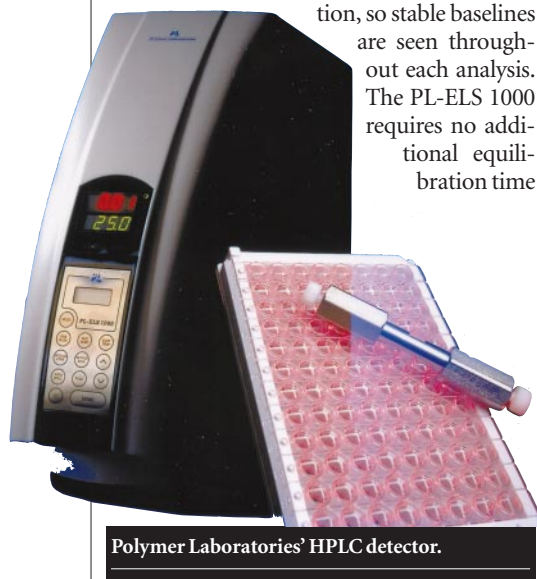
Reader Service No. 100

PL-ELS 1000

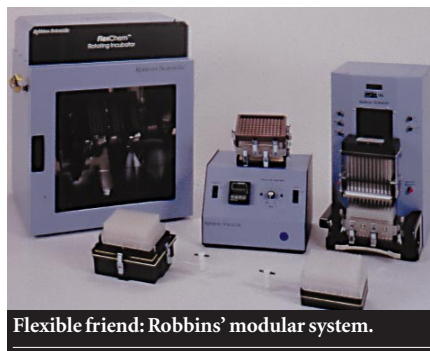
Polymer Laboratories www.polymerlabs.com

An HPLC detector for HTS applications

This evaporative light scattering detector for high-performance liquid chromatography (HPLC) is recommended for high-throughput screening of combinatorial chemistry libraries. The PL-ELS 1000 can detect compounds with no UV chromophore and permits rapid separations using extreme gradients. The detector's baseline does not respond to changes in eluent composition, so stable baselines are seen throughout each analysis. The PL-ELS 1000 requires no additional equilibration time



Polymer Laboratories' HPLC detector.



Flexible friend: Robbins' modular system.

for the signal to return to the starting point, thus increasing the high-throughput capabilities of the detector.

Reader Service No. 101

FlexChem

Robbins Scientific

www.robsci.com

High-throughput synthesis

A modular system for performing high-throughput, parallel synthesis of small molecules, FlexChem consists of equipment and disposables specially designed to streamline and accelerate the preparation of compound libraries. The heart of the system is a disposable, chemically resistant 96-chamber reactor block with reusable, independently removable sealing covers that attach to the reactor block to simplify the addition of resin and reagents without leaking. Up to 384 reactions can be performed simultaneously using the FlexChem Rotating Oven, which provides temperature control and agitation of four FlexChem sealed reactor blocks. The Hydra-96 FlexChem dispenser can add resin and solvents to all 96 chambers of the blocks simultaneously.

Reader Service No. 102

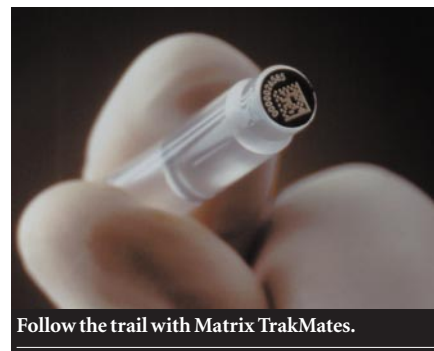
TrakMates

Matrix Technologies Corporation

www.matrixtechcorp.com

Keeping track

This new line of storage tubes is, claims Matrix, the industry's first to feature 2-D Data Matrix Codes. Available with a capacity of 1.4 ml, TrakMates tubes are marked with up to 10 billion unique codes. This identification system ensures a totally secure audit trail from compound to hit, even on the large HTS projects. Each code is laser-etched to resist mechanical and chemical damage and, if a flaw occurs, an Error Correction Code (ECC 200) configuration allows the code to be read even if as much as 25% is damaged. TrakMates tubes feature dual high-contrast white-on-black labelling. They also have a human-readable code to



Follow the trail with Matrix TrakMates.

double-check accuracy and simplify identification in locations without scanners.

Reader Service No. 103

Peptide library

Sigma RBI

www.sigma-aldrich.com

Pharmacologically active peptides off the shelf

Sigma-RBI's LOPAP (Library of Pharmacologically Active Peptides) includes 60 pharmacologically active peptides with demonstrated agonist activity. They are well characterized, high-purity peptides which can be used to characterize orphan receptors, validate new HTS assays and help to define pharmaceutically relevant screening strategies. LOPAP is organized in a 96-well plate format. Each well contains a measured quantity of peptide to give a 1 mM concentration upon the addition of 100 µl of solvent. LOPAP also includes a plate map of the peptide positions, the peptide name and amino-acid sequences, Sigma-RBI catalogue numbers and a diskette containing an SD file (Mac and PC formats) for use with ISIS Base or other compatible applications.

Reader Service No. 104

TopCount NXT

Packard Instrument Company

www.packardinst.com

Making use of small-volume microplates

Enhancements to the TopCount NXT microplate scintillation and luminescence counter make it suitable for HTS applications. Packard's new shallow-well microplates increase counting efficiency for scintillation proximity assays, and help reduce reagent consumption to lower the overall cost of these assays. The plates have small-volume wells, which place the samples closer to the detectors and reduce the effect of colour quenching. Improved performance is particularly noticeable for 3H SPA PVT assays, which have up to 70% higher counting efficiency compared with deep-well microplates. The recent

new on the market

revision of the TopCount NXT software, combined with a faster internal computer, greatly enhances the instrument's performance.

Reader Service No. 105

Biomek FX

Beckman Coulter www.beckmancoulter.com

Liquid handling for HTS

The Biomek FX High-Throughput Screening system aims to simplify the process of creating a method and increases the overall throughput of the SAGIAN Core System. With up to two liquid handling arms and a flexible, high-capacity work deck available, the Biomek FX integrates multiple liquid handling capabilities with 96- and 384-channel heads and an independent eight-channel pipettor. Available with single or dual interchangeable pipetting heads, the Biomek FX allows the use of one or two pipettors. The 96-channel head is now available. Eight- and 384-channel heads for this platform were previewed in January.

Reader Service No. 106

[FP]² assay technology

NEN www.nen.com

Non-radioactive screening of G protein-coupled receptor targets

The [FP]² family of homogeneous fluorescence polarization binding assays provides HTS labs with a simple, robust, automation friendly, non-radioactive solution for screening GPCR (G protein-coupled receptor) targets. Proven for 96- and 384-well formats, and extendable to 1,536-well, the [FP]² family of assays uses a homogeneous 'mix and measure' protocol. The assays have been demonstrated with commercial FP readers and a broad range of cloned GPCR membranes, while providing fast count times (less than 1 second per well).

Reader Service No. 107

Purified protein

Corning www.corning.com

Microplates for HTS assays

If you need to increase your HTS assay productivity as well as achieve greater consistency and accuracy, then Corning's new 96-well coated microplates should interest you. Available in a choice of clear, white or black, and pre-blocked ready for use, the microplates are coated with streptavidin, a 60KD protein isolated from *Streptomyces avidinii*. The purified protein is bound to the wells of the polystyrene microplates using a proprietary coating technique unique to Corning. The company claims that the technology ensures a high binding capacity of biotin together with high coating homogeneity, low leaching of streptavidin (<6 ng per well) and high resistance to commonly used detergents. The uniform surface is totally stable and its high specificity improves the signal-to-noise ratio.

Reader Service No. 108



Somewhere for your cells: the new assay plates from BD Biosciences.

Biocoat

BD Biosciences www.bd.com

New plates for a variety of assays

BD BioCoat Collagen I and Poly-D-Lysine (PDL) 96- and 384-well assay plates are now available from BD Biosciences for use in a variety of cell-based fluorescence, luminescence, calorimetric and radiometric assays. This ready-to-use format offers reliable performance and lot-to-lot consistency. BD BioCoat Collagen I and PDL Cellware is claimed to improve cell adhesion in a variety of normal and transfected cell lines. The plates are available in clear, white and black opaque with clear bottoms, and solid white opaque formats.

Reader Service No. 109

Chirobiotic kit

Astec www.astecusa.com

Enantiomeric separations

This new kit includes three Chirobiotic phases developed for the separation of enantiomers based on bonding amphoteric macrocyclic glycopeptides to a 5 µm silica gel. Chirobiotic V is bonded with vancomycin, Chirobiotic T is bonded with teicoplanin and Chirobiotic R is bonded with ristocetin A. The result is a range of broad-based columns that can be run in three different solvent modes making them applicable for a wide variety of enantiomeric separations. Three basic mobile phase systems are used together with column coupling to achieve very broad chiral selectivity in the shortest possible time. Optimization techniques are also simplified since an Rs of 0.6 or greater on a 10 cm Chirobiotic column coupling screen can be optimized to Rs >1.5 on a 25 cm column for the selected stationary phases. The kit contains one of each of Chirobiotic 100 × 4.6 mm columns, two column couplers, a Chirobiotic handbook and a Method Development Kit Operating Manual.

Customer Service No. 110

FLIPR Calcium Assay Kit

Molecular Devices www.moleculardevices.com

Reagent kits for calcium assay

This new assay is the first in a series of Molecular Devices reagent kits targeted at high-throughput screening and drug discovery applications. The FLIPR Calcium Assay Kit was developed to improve the screening efficiency of the FLIPR system, a drug discovery platform used by all of the largest pharmaceutical companies. The new kit provides all the reagents needed for the predominant screening application performed on the FLIPR system, the calcium assay. The kit's unique feature is that it enables researchers to eliminate a step in the FLIPR calcium assay protocol, thereby saving up to 15 minutes of processing time for each 384-well plate. This kit has the potential to increase throughput, reduce costs and improve screening efficiency.

Reader Service No. 111

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

Look no further
for the best value in
Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA
GENOSYS (+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33

The kiss of life?

Chemistry rediscovers itself as basic research gets a boost

p807

All change!

The European chemical industry is radically changing its R&D policies

p809

Out of favour

Across Europe, fewer students are electing to study chemistry

p809

All together now

Europe plans networks to enhance basic chemical research

p811

'Quiet revolution' in chemistry could revive public and private sectors

New frontiers in basic chemistry could drive applied science in many fields. But support for basic research seems to be waning, in both academia and industry. Paul Smaglik dons his lab coat.

Physicists have black holes to discern. Biologists have genomes and proteomes to decipher. Chemistry, with its periodic table firmly in place, would appear to be short of grand goals and new frontiers. But that perception could not be further from the truth, says Stephen Lippard, head of the chemistry department at the Massachusetts Institute of Technology (MIT).

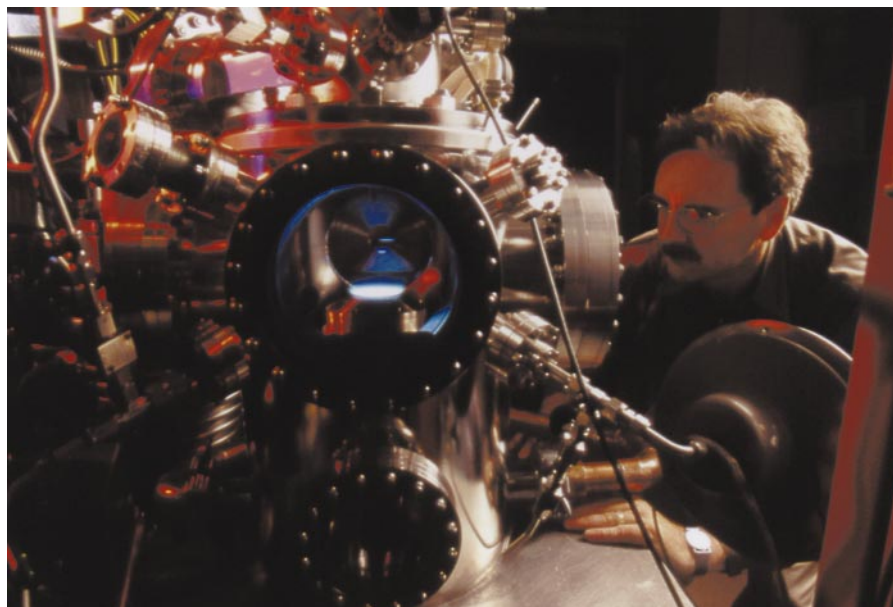
Lippard polled colleagues involved in chemistry from around the United States about areas ripe for discovery. He compiled the responses into a list of 22 new frontiers in basic chemistry (see *New Frontiers in Basic Chemistry*, overleaf). Work towards those goals is causing a "quiet revolution" in chemistry, Lippard says.

Turning up the volume could lift the fortunes of both academic and industrial chemistry. Both sectors need that boost. Basic research in industrial chemistry occupies a shrinking percentage of companies' already tightening budgets. And basic research in academic chemistry is being overshadowed by the continued growth in federal support for biology.

Lippard notes that other fields sometimes take basic chemistry for granted. This is unfortunate, he says, because it serves as both an enabling technology — it helped make high-throughput sequencing of the human genome possible, for example — and a common language that scientists in many disciplines use to communicate. Investing more in basic chemistry will advance other fields, and help the language of science evolve, he believes.

Academic road-map

Lippard's wish-list has received a lot of support. "These fundamental milestones are worthy of pursuit in and of themselves, but by meeting them, they will also drive other fields, such as biology, medicine, materials, agriculture, biotechnology and nanotechnology," agrees Peter Dervan, a



Into the future: fuel cell research at ExxonMobil demonstrates the application of basic chemistry.

professor of chemistry at the California Institute of Technology. He adds that Lippard's report is especially valuable for the university administrators who allocate resources.

George Whitesides, a chemistry professor at Harvard University, notes that the list will also help non-chemists — perhaps policy-makers who make the decisions about federal research funding. "If you're an outsider, you don't know what a given field is doing," he says.

But communicating the need for funding in these areas is not enough, he adds. The community must take on the problems and then demonstrate that these issues are "sufficiently interesting and sufficiently useful" if they wish to see them funded further.

For academic chemists seeking funding from both industry and the federal government, the take-home message from Lippard's report might be to conduct basic research, but with an eye towards practical

applications, says Craig Hill, professor of chemistry at Emory University in Atlanta, Georgia. Hill has used that approach to fund his work in understanding and improving catalysis, one of Lippard's 22 goals.

Hill originally envisioned his research on creating a catalyst that assembles itself in water as a basic science project. It was first funded by the National Science Foundation as such, but as the work advanced, the implications that such a catalyst could result in cleaner chemical processes caught the attention, and funding, of the paper-making industry.

A basic approach

Nicholas Turro, professor of chemistry at Columbia University in New York, says that Lippard's list is helpful for students, because it identifies exciting problems. Young scientists wanting to get grants should approach basic problems in chemistry with an eye on their potential

EXXONMOBIL

applications. "If I'm going to do something that's intellectually pure, why not try to overlap it with something on this list that's applied?" he says.

That may be easier to do for some items on Lippard's list than for others, notes Patricia Watson, DuPont's manager for chemical sciences and engineering research. For example, the company is actively working on Lippard's challenge to explore chemistry at interfaces, because understanding this could lead to paints or coatings that reduce or eliminate corrosion. But many of the challenges, such as creating self-replicating molecules, have less obvious applications. "That's fancy and interesting," Watson says. "But we're not doing that."

In general, many of Lippard's goals call

for designing specific molecules. But many sectors of the industry are more interested in screening thousands of existing compounds. "We would take a more empirical approach, rather than trying to build things from the ground up," Watson says.

She also notes that companies have to pick and choose the areas where they conduct basic research carefully. Watson estimates that DuPont does less basic research now than it did ten years ago — largely because of pressure to keep costs down. That pressure will only increase as chemical companies continue to merge. For example, when Exxon combined with Mobil in December 1999, the new company announced it would slash about 15% of its work force, some 16,000 jobs, by the

end of 2002. This month, the company decided to cut an additional 3,000 jobs. Although it is unclear how many of those positions are in research, basic science is sure to be affected.

Despite this, Watson believes basic research will never be completely eliminated from chemical companies' agendas. "Industry still needs proprietary positions," she says. "There will always be basic research in key targeted areas."

A question of maturity

Although she agrees that the list identifies fertile areas of activity, Watson does not quite accept Lippard's underlying thesis that chemistry is not a mature science. The challenges on his list focus on more sophisticated, specific ways to manipulate molecules, not on outlining a paradigm shift for the field, she says. "The revolution is not just in chemistry. It is in the marriage between all the disciplines," she adds.

Making that marriage work means that chemists need to be more flexible, have deeper expertise in their speciality, and a broader understanding of how their expertise fits in with other disciplines, says Jose Santiestaban, section head for ExxonMobil. "The chemist of the future has to have a certain expertise, but they also have to speak the language of the other fields," he says.

Interdisciplinary literacy will help the revolution in basic chemistry spill over into the pharmaceutical industry, says Malcolm MacCoss, a Merck vice-president for medicinal chemistry. He estimates that transforming a university-trained synthetic chemist into a productive, industrial-based medicinal chemist takes seven to eight years. The position demands a firm grasp of basic chemistry, but success increasingly comes from mastering subdisciplines, such as computational biology.

This kind of expertise is needed to pursue those of Lippard's goals that are especially applicable to the pharmaceutical industry, notably developing many variations of a molecule that bind to a specific target. "You'd like a compound that is potent and exquisitely selective for the target that you want the molecule to interact with," says John Kozarich, Merck vice-president for biochemistry and basic research.

But designing such a dream molecule can result in clashes between basic chemistry's elegance and applied chemistry's reality. Designing a molecule with many properties often results in a bigger molecule. Bigger molecules tend to be more expensive to manufacture and less bioavailable, which makes oral administration less feasible. So work at the new frontiers may lead to a quiet revolution in chemistry — but not an easy one. ■

Paul Smaglik is *Nature's* correspondent in Washington DC and Careers and Recruitment editor.

New frontiers in basic chemistry

- Create chemical entities made up of many identical components arranged in a way as to serve as receptors for given binding units or ligands.
- Create self-replicating molecules and self-correcting chemical reactions. Included in this latter category are catalytic reactions where errors introduced in the product are deleted and corrected.
- Understand the nature of a given or desired new chemical transformation sufficiently well that a catalyst for the reaction can be designed in a rational manner.
- Explore the chemistry at interfaces and control the stereochemistry of heterogeneous catalysts used as bulk species.
- Harness supramolecular constructs to preserve the integrity of chemically labile species while conserving sites for the catalysis of chemical reactions including product release.
- Use parallel, automated synthesis or combinatorial trial-and-error reactions to improve on a chemical transformation in an evolutionary manner.
- Control the direction and orientation of the approach of one molecule reacting with another.
- Understand the internal motions of molecules so that, with appropriate timing, a pulse of electromagnetic energy can be used to dissociate specifically a chosen bond in the molecule.
- Understand the structure and dynamics of intermolecular interactions so that the properties of molecular collections can be predicted and controlled.
- Devise reagents and pathways for breaking into chemical bonds previously considered to be inert.
- Find the means for converting naturally abundant substances into chemically useful small-molecule building blocks and vice versa.
- Develop the art of conducting chemical reactions without solvents.
- Create reagents that chemically modify a portion of a molecule without protecting, and later having to deprotect, other reactive sites on the molecule that we do not want to modify.
- Create chemical products and processes that do not require the use or generation of hazardous substances and that use renewable resources.
- Understand and harness the properties of compounds of intermediate (1–100 nm) size between the molecular and solid state.
- Investigate the chemistry of single isolated molecules and compare the properties with those of an ensemble of such molecules and in various solvents.
- Create molecules that self-assemble into supramolecular structures including solids, the properties of which surpass those of the molecular collection.
- Learn how to grow crystalline solids of molecules of any composition and to incorporate guests into these crystals to enhance their physical and chemical properties.
- Discover unusual compositions of matter comprising chemical elements and fundamental particles in combinations not previously encountered.
- Master the chemistry of caged species, releasing at will a guest entrapped in an appropriate host in the gas, solution, or solid state by introducing a chemical, magnetic or electric field.
- Understand and use the chemistry, energetics and spectroscopy of polyatomic radicals.
- Evolve new theoretical approaches to understand chemical bonding and reactions and to test these theories against real chemical systems.

European industry turns to the academics to secure its future

In Europe, the days of the chemical industry's great company laboratories, with their rich tradition of fundamental research, are all but gone. For nigh on a hundred years they were key players in basic chemistry, and industry-based research collected accolades and Nobel prizes. Driven by economic necessity and the restructuring following global consolidation, most firms have shifted their research focus so that it is now firmly product-orientated and much narrower in scope than in the past.

"I have a sense of long-term innovation moving away from industry," says André Merbach, professor of chemistry at the Swiss University of Lausanne and a prime mover in efforts to find new ways for industry and academia to learn of one another's needs. "Restructuring is leading to business units with a much narrower focus interested mainly in product-orientated research. I have a sense, too, of their being fewer researcher posts in industry and more collaboration between the big companies and start-ups around universities."

Although the phenomenon is too new to have been turned into a set of comprehensive and definitive pan-European statistics, Merbach's comments ring true, and are supported by observations made by academics and

industrialists from all corners of Europe. Slowly, much of the chemical industry's in-house research is becoming short term.

Ironically, the chemical industry's commitment to undertaking its own basic chemistry is waning at a time when the field is vibrant. Chemists are gaining a deeper understanding of how individual atoms and bonds behave. They also have better computational methods for modelling molecular interactions. They are experimenting with new materials with properties that are generated not by changing the molecules involved, but by the way these molecules are organized and arranged within the material.

A centralized start

Until the 1990s, basic research was industrialized in line with the model that emerged at the end of the nineteenth century. In his book *The Norton History of Science* (Norton, 1992), William Brock, a science historian at the University of Leicester, credits Carl Duisberg (1861–1935) from the German company Bayer with transforming what were then the chemical industry's piecemeal research efforts. Duisberg's work on azo dyes earned money for Bayer. Impressed, the company allowed him to reorganize its laboratories.



Industry needs links with academic research.

He opened a new central laboratory for research chemists, and provided the researchers with an extensive library and an in-house abstracting service for more than 500 journals and magazines.

An experimental dye-house evaluated the ideas emerging from the basic research. Analysts studied potential products and those from competitors. Chemical engineers scaled up ideas. Specialists studied the patents of rivals and patented their own ideas. "It was the pattern of industrialized invention," writes Brock, "with its potential

Undergraduate interest in chemistry wanes in Europe

Across Europe, chemistry is losing its appeal as a subject to study. This is part of a declining interest in physics, chemistry and biology, says Jan de Wit who is a professor of research management at the University of Nijmegen and staff director for research at Akzo Nobel.

"In Italy," says Claudio Bianchini head of the Florence-based Institute for the Study of Stereochemistry and Energetics of Coordination Compounds, "the number studying chemistry is decreasing dramatically." Peter Noordervliet, director of the Association of the Dutch Chemical Industry (VNCI), adds: "There are two-thirds fewer chemistry undergraduates in the Netherlands today than there were in 1990. For the first time last year, the demand for chemists exceeded supply."

In Germany too, fewer students are studying chemistry, says Dieter Jahn, senior vice-president for university relations and research planning at BASF. "This is not yet a problem for the

company," he says, "but it is not a good thing and there will be fewer qualified chemists in future years."

The decline is attributed by de Wit to the comparative difficulty of what he calls "the exact sciences" compared with other subjects, and to the perception that chemists will end up poorly paid and working for people who studied economics at university. "They think if that's the case, they may as well study economics themselves," de Wit says.

When he lectures undergraduates, de Wit does his best to dispel the attitude that scientists have few career options. He tells them that when a chemist joins industry there are plenty of opportunities to move into marketing, business development or patenting. "By the time you reach 40, if you are still in the lab, that is probably because you want to be," he says.

"At BASF," says Jahn, "90% of people who joined the laboratory will know what direction, say marketing or business development, they want to

take their careers." However an increasing minority stays in the laboratory to help preserve the memory of institutional research, he says.

Bianchini attributes some of the fall in numbers of chemistry students to a mixture of negative public perception of chemistry and to the subject's difficulty. "When I tell people in Europe I am a chemist, they look at me as though I should not exist," he says.

For those graduates and postgraduates brave enough to withstand disapproving looks, the shortage of chemists is good news. "Only a few years ago my research students found it very hard to get a job. Now it is unusual to see a good chemist without a job," says Bianchini.

The shortage may be good news for individuals, but there are more serious consequences. Bianchini and de Wit say it is hard to find good researchers. The problem, says Noordervliet, will get worse in the Netherlands before it gets better. Even if there is an influx of eager chemists,

it will take seven years before they are qualified researchers. In the meantime, job hunters are attracted by the higher wages of industry compared with academia. "The problem is worse for academia than industry," says Noordervliet. "Students from Romania, Poland and Russia are taking the posts, but these people will go back. We are losing our science base."

Faced by this situation, Noordervliet's association is working to improve chemistry's image. For example, the VNCI funds a theatre group that tours schools telling children as young as seven or eight about the role of chemistry in the kitchen.

At a European level, the European Chemical Industry Council is working with 16 European science and technology museums to provide a website called *Chemistry for Life*. Visitors to the site can participate in a variety of exercises which give them an overview of chemistry in everyday life.

http://www.chemforlife.org

H.G.

► for expanding the scientific workforce, that became the international model for twentieth century industry."

A hundred years later, Bayer is still one of the world's leading chemical company (see table). But the central research laboratories that it pioneered have lost their lustre. Globalization and the sheer breadth and complexity of the chemical industry has changed all that.

"When I joined Akzo Nobel in 1973, industrial research was knowledge-driven," says Jan de Wit, staff director for technology at Akzo Nobel in the Netherlands and a professor of research management at the University of Nijmegen. "Then we could say we had the best people in the field, it had good PR value, but we did not do much business. The research could have a 10- to 15-year lead time," he says. Now the maximum time for

which industry can do research before having a product is five years, he says.

Bucking the trend

There are exceptions. The German company BASF, for example, spends 10% of its R&D budget on what it calls exploratory research. Although this is not strictly basic chemistry in the sense of being entirely about the generation of new theoretical ideas, neither is it entirely product-orientated. The exploratory research can have a lead time of ten years, and will examine concepts that could underpin several different branches of the company's work. "For example, some 85% of our processes are catalytic," says Dieter Jahn, senior vice-president for university research and research planning, "be they for plastics, chemicals, fibres or vitamins."

The company pays centrally for the exploratory research. The rest of the nearly 1 billion euros that BASF spends on R&D (excluding pharmaceuticals) is product-orientated and spent by the business unit. The laboratories feeding products to the business units are all located together, and Jahn is conscious of the potential this gives for cross-fertilization. "My experience is that it is at the interface between research areas that discoveries are most intensive," he says.

Advances welcomed

Despite the shift away from basic research, academics argue that Europe's chemical industry needs fundamental advances to remain competitive. Although the European Union's chemical industry leads the world in terms of sales (at 30%), hosts the corporate headquarters of 17 of the top 30 chemical companies and employs roughly 1.7 million people, the sector struggles to be as profitable as that of the United States. And, despite more than a decade of work to improve its environmental record, the industry suffers from a relatively negative public perception (see box, left).

Research breakthroughs are key to a green-

Table 1 Top 15 global chemical companies, 1998

Company	Sales, billion US\$*	Region
Bayer	31.2	Europe
BASF	30.7	Europe
Merck	26.9	US
Hoechst	24.8	Europe
DuPont	24.8	US
Novartis	21.9	Europe
Dow	18.4	US
Roche	17.0	Europe
ICI	15.5	Europe
Rhône-Poulenc	14.7	Europe
Akzo Nobel	13.9	Europe
Glaxo Wellcome	13.3	Europe
Pfizer	12.7	US
Shell	12.3	Europe
Henkel	12.1	Europe

*Averaged exchange rate for the year; Source: Chemical Insight/CEFIC

er future. "Only basic chemistry can provide the insights that will lead to more environmentally friendly processes," says Claudio Bianchini, director of the Institute for the Study of Stereochemistry and Energetics of Coordination Compounds (ISSECC) in Florence, Italy. "We need less toxic methods of synthesis and more selective processes that create fewer waste byproducts."

Now that the dust is settling on the closure of its central research facilities, industry itself sees that it needs to find new ways to relate to basic chemistry. "The most dramatic change in the past few years has been industry's realization that it needs the academics," says David Bricknell, responsible for science and technology policy at the Brussels-based European Chemical Industry Council.

A number of schemes exist to bring basic and industrial researchers together. The Netherlands, for example, has established a centre of excellence for polymer research at the Eindhoven University of Technology. Polymer research is a hot topic for both industry and academia, and the centre pools expertise from different Dutch universities and companies.

When de Wit is wearing his academic hat,

Chemistry and the environment

The European public remains critical of the environmental and health risks posed by chemicals, and disapproves of the chemical industry's business practices and approach to social responsibilities, according to a survey conducted by the European Chemical Industry Council (CEFIC).

CEFIC surveys public opinion every two years. The Pan-European Survey 2000 is based on interviews with 9,000 citizens from nine countries. Positive attitudes towards the chemical industry vary from a high of 59% in Germany to a low of 24% in Sweden. On average, 45% have a positive attitude to the chemical industry compared with 42% in 1998.

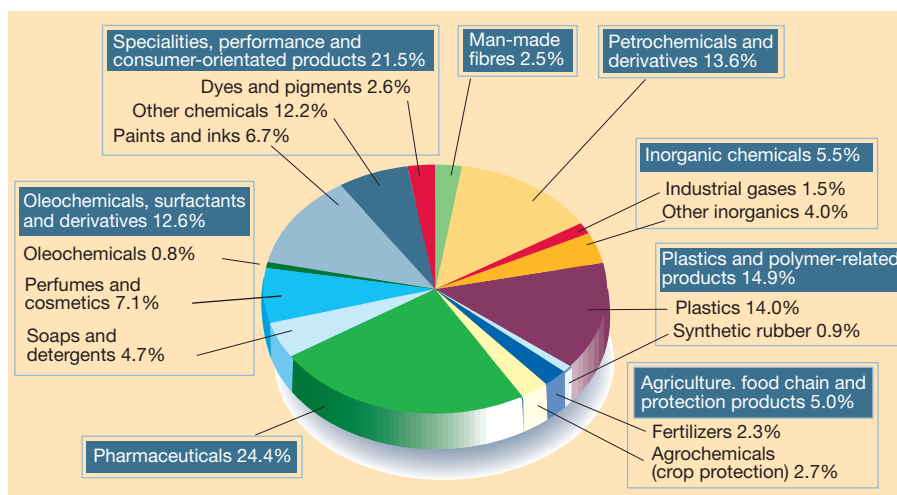
Although 54% thought that the chemical industry provides "tangible benefits to the quality of life through new and safer products", the industry's reputation lags behind those of electronics (82%), the electricity industry (77%), food industry (76%), pharmaceuticals (72%) and the car industry (70%).

"The image of chemistry with the public is unfortunately not very good; it is going to be difficult to reverse, but we must if we are to have bright young people in our labs and running our plants," says Jean Gauthier-Lafaye, director of research at the French-based company Rhodia.

"The biggest problem the chemical industry faces," says David Bricknell, responsible for science and technology policy at CEFIC, "is the public's ignorance of the role of chemistry in our daily lives and the way the chemical industry underpins our standard of living."

Bricknell argues that the industry has done a lot to clean up its act during the past ten years. This may be true, but Claudio Bianchini, a professor of chemistry at Italy's National Research Council, says: "Many industrial processes are [still] old in terms of efficiency and environmental compatibility."

It seems that chemistry in Europe faces two problems. It must continue to improve industrial processes, and it must win public confidence. **H.G.**



Division of labour: the various sectors of the European chemical industry by production for 1998.

which he does one day a week, he researches how best to bridge the gap between the ideas that academics generate and the product development industry seeks. At Akzo Nobel, where the bulk of R&D spending goes on product development, de Wit encourages the 23 business unit managers to identify their research needs and arranges contracts between them and the universities. He is keen to explore the intellectual output of academic chemistry in countries such as Russia, where western companies have in the past found it less easy to learn what was happening.

Bianchini, a professor of chemistry at Italy's National Research Council (CNR) as well as the director of ISSECC, sees the situation from an academic's viewpoint. "During the past five years we have received an increasing number of proposals to collaborate with industry," he says. "In the past, industry only wanted us to look at ways to improve known processes. Now we have more freedom to do fundamental research. We are surviving with industry money."

Trades and trade-offs

The advantages to industry, says Bianchini, is that this approach is less expensive and they have access to a deep pool of academic

talent. If they employed their own basic researchers, the same people would stay with the company.

The story is the same in the Netherlands, says Peter Noordervliet, director of the Association of the Dutch Chemical Industry. Industry pays for more than half of the chemistry research in Dutch universities compared with only a fraction of this a few years ago.

But academics pay a price for this industrial largesse, according to Bianchini. "Industry patents every result, even if it is a long way from commercialization. This leads sometimes to a two- or three-year delay in publication. This may not be a problem for professors, but it is for young researchers," he says.

Given the pressures facing the European chemical industry, its desire to keep its global position and its importance as a major employer, the subject of how industry and academia can best relate and resolve issues such as intellectual property rights is likely to be a focus of attention for national policy-makers throughout Europe.

Helen Gavaghan is a science and technology writer based in Hebden Bridge, West Yorkshire.

VNCI ► <http://www.vnci.nl>

Akzo Nobel ► <http://www.akzonobel.com>

CNR ► <http://www.cnr.it>

would benefit from interaction with others, it seeks approval from the COST technical committee. This makes COST a good indicator of emerging fields in basic research.

For example, Antonio Laganà, a professor of computational chemistry at the University of Perugia in Italy, is trying to establish a research network in what he calls meta-computing. Some chemistry research problems and simulations are so complex, he argues, that they cannot be undertaken by one research institution alone, even if it is a centre of excellence. Laganà's idea fits the current thinking of Philippe Busquin, the European Commissioner for research, that Europe should develop virtual laboratories linking research facilities across Europe.

Laganà does not expect the meta-computing network to be adopted by COST before the end of the year, but he already has three proposals for evaluation. Two are for education and would enable laboratories to share software and access computers with spare capacity. David Bricknell, who is responsible for science and technology policy at the European Chemical Industry Council, sees potential for industry if computers are linked. "Massive parallel computing might allow us to design and operate processes differently," he says.

One COST action that is up and running is chaired by Bianchini. It aims to promote collaboration between groups researching the synthesis of oligomers, polymers and copolymers via metal catalysis. The field is attracting enormous attention at a basic research level because the scientists are exploring new approaches to modifying the properties of materials. Instead of changing the chemistry to obtain new properties, they are manipulating the physical positioning, orientation and arrangement of molecules.

The field is also important to industry. Jan de Wit, staff director for technology at Akzo Nobel in the Netherlands, says: "We are discussing with universities how to make catalysis sustainable. In the past if a catalyst became unusable because it was covered by

Emerging fields of basic chemistry in Europe

Driven by improved instruments that can tease out fine spectral and spatial details, and by a deepening understanding of how matter behaves at the level of individual molecules and bonds, fundamental chemistry is passing through a revolutionary phase. This is the "quiet revolution" described by Stephen Lippard, chairman of chemistry at the Massachusetts Institute of Technology (see page 807), an event that Claudio Bianchini, a professor of chemistry at Italy's National Research Council (CNR), wholeheartedly agrees is occurring.

In Europe this revolutionary path can be seen by looking at the growing number of research networks that form part of a venture known as the European Cooperation in the field of Scientific and Technical Research (COST). This is essentially a networking exercise intended to coordinate fundamental research with funding from national bodies, says Brian Hanley from the Brussels-based secretariat. At present, 33 countries from the European Union, central and eastern Europe participate. One focus for COST is chemistry.

Funds come from the European Union's research budget and are spent on conferences, travel grants and short-term grants allowing young scientists to visit a collaborating group for a few weeks. COST also pays

for short-term projects such as the collation of a valuable contact list of the names, research interests and contact details of all European chemists participating in COST chemistry research networks.

Unlike other research efforts backed by the European Union, COST is driven by the scientists themselves and is not undertaken in response to requests for proposals. If a group thinks that a particular branch of chemistry — combinatorial chemistry, for example —

Who's who in Europe

The two starting points for chemistry in Europe are CEFIC — European Chemical Industry Council (<http://www.cefic.be>)

and the European Network for Chemistry (<http://www.chemsoc.org/networks/enc/>)

CEFIC provides links to all the national associations for the chemical industry in Europe as well as extensive facts and figures about the industry and position papers on significant policy issues.

AllChemE (<http://www.chemsoc.org/networks/enc/AllChemE.htm>) is an alliance of five organizations concerned with academic and industrial chemistry in Europe, education and training, and policy

discussions. As well as CEFIC, they are:

CERC3, the Chairmen of the European Research Councils Chemistry Committees

(<http://www.chemsoc.org/networks/enc/cerc3.htm>)

COST, European Cooperation in the Field of Science and Technical Research

(<http://www.unil.ch/cost/chem>)

ECCC/FECS, the European Communities Chemistry Council/Federation of European Chemical Societies

(<http://www.chemsoc.org/networks/enc/eccc.htm>);

(<http://www.chemsoc.org/networks/enc/fecs.htm>)

EFCE, the European Federation of Chemical Engineering

(<http://www.dechema.de/efce.htm>)

► layers of byproduct, we left the customer to dispose of the catalyst as waste. Now we take it back and clean it up for resale.”

Although Bricknell's and de Wit's observations show that meta-computing and new approaches to research on polymerisation via metal catalysis might eventually be of value to industry, the ideas that scientists are exploring in each field are still in their infancy.

All the same, the organizers of COST networks and of industrial networks promoting the sharing of best practice have realized that there is potential overlap between what the universities are doing and what industry

wants. “About 18 months ago,” says Hanley, “the COST chemistry secretariat began to have informal discussions with an industry network called SUSTECH.”

SUSTECH (short for sustainable technology) was established by a group of European companies in 1994. SUSTECH and COST will soon publish a report listing fields where joint working groups will be formed to promote collaboration between industry and academia. Polymerization via metal catalysis and meta-computing for chemistry are likely to be fields of joint action. **H.G.**

COST ♦ <http://www.unil.ch/COST/chem>

US companies seek basic science



Cleaning up: paper making could be made a cleaner process with new catalysts.

Several goals for basic chemistry are especially attractive for industry. These include increasing the ability to manipulate chemical bonds, developing more efficient catalysts, and lessening industry's dependence on solvents. Meeting these goals could reduce pollution and save money.

“The chemical industry is aiming for zero waste, zero emissions,” says Jose Santiestaban, section head for ExxonMobil.

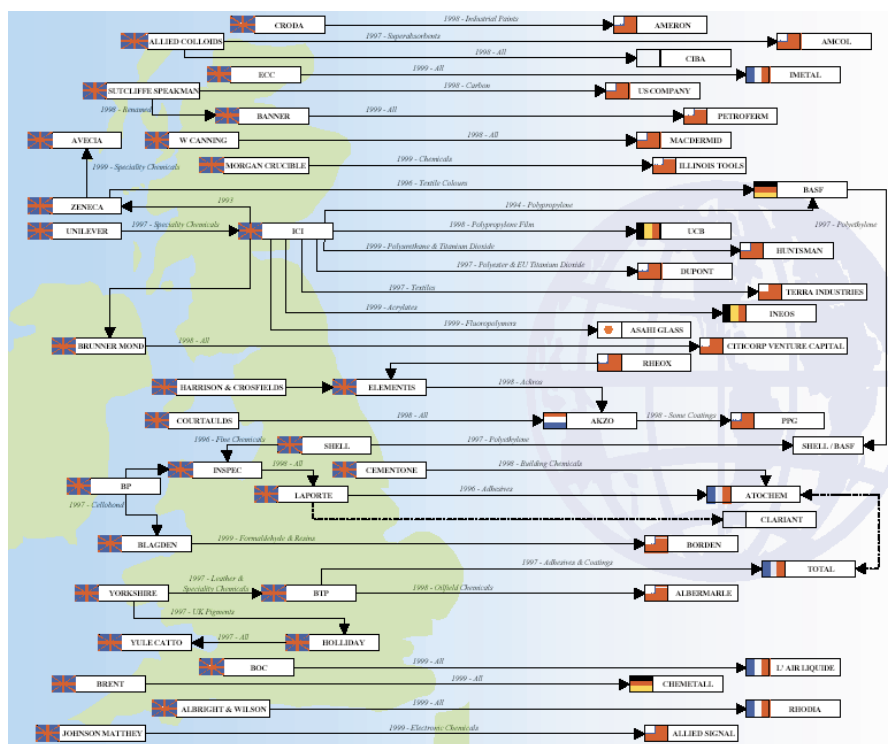
Santiestaban says his company is interested in turning methane (CH_4) — often generated as a waste product in petrochemical production — into high-value molecules such as polyethylene or polypropylene. Because it is so unreactive, methane needs to be activated in some way to make it useful as a feedstock for a reaction. Increasingly sophisticated knowledge of the bonds between carbon and hydrogen atoms is bringing this goal closer.

The paper and pulp industry is another potential beneficiary from basic research. Craig Hill, professor of chemistry at Emory University in Atlanta, Georgia, is working on a catalytic system that could make paper manufacturing a cleaner enterprise. The centrepiece of his technique is a mixture of metals in water that, at the right temperature, ‘self-assemble’ into a catalyst.

Catalysts almost always degrade during chemical reactions, often because of the temperatures they have to work at, and this results in waste. So Hill is seeking a system that is stable under the reaction conditions.

In addition, Hill's catalyst system is water-based, which eliminates the need for environmentally unfriendly solvents. The catalyst could be used to remove lignin from wood pulp to leave just the cellulose, which can then be turned into paper. Usually, paper and pulp companies need to use organic solvents, which can often end up in rivers. **P.S.**

Mergers and acquisitions rock UK chemical industry infrastructure



A mass of mergers: the fate of UK chemical companies in the 1990s.

The global consolidation of chemical companies has reduced the number of laboratories and caused a decline in industrial research and development, says Britain's Royal Society of Chemistry (RSC) in a report published earlier this year.

In the UK, money spent by the chemical industry on R&D went down by 16% between 1993 and 1997, according to the report. In addition, total UK employment in chemicals R&D fell by 31%, with the biggest job losses among scientists and technicians.

The UK came out of the consolidation of the 1990s badly, and non-UK companies now own many famous British names. Akzo Nobel, for example, took over Courtauld's,

and Albright & Wilson is now part of the Rhodia groups of companies (see above).

“It looks as though the UK's chemical industry might go the way of the car industry,” says David Bricknell of the European Chemical Industry Council.

As with the rest of Europe, the Royal Society of Chemistry found that mergers, acquisitions and restructuring had resulted in the closure of corporate research laboratories in the United Kingdom. And universities reported concerns that the reorganization had put a blight on planning research, which could affect the long-term relationship between industry and academia. **H.G.**

♦ <http://www.rsc.org/pdf/general/m&aureport.pdf>